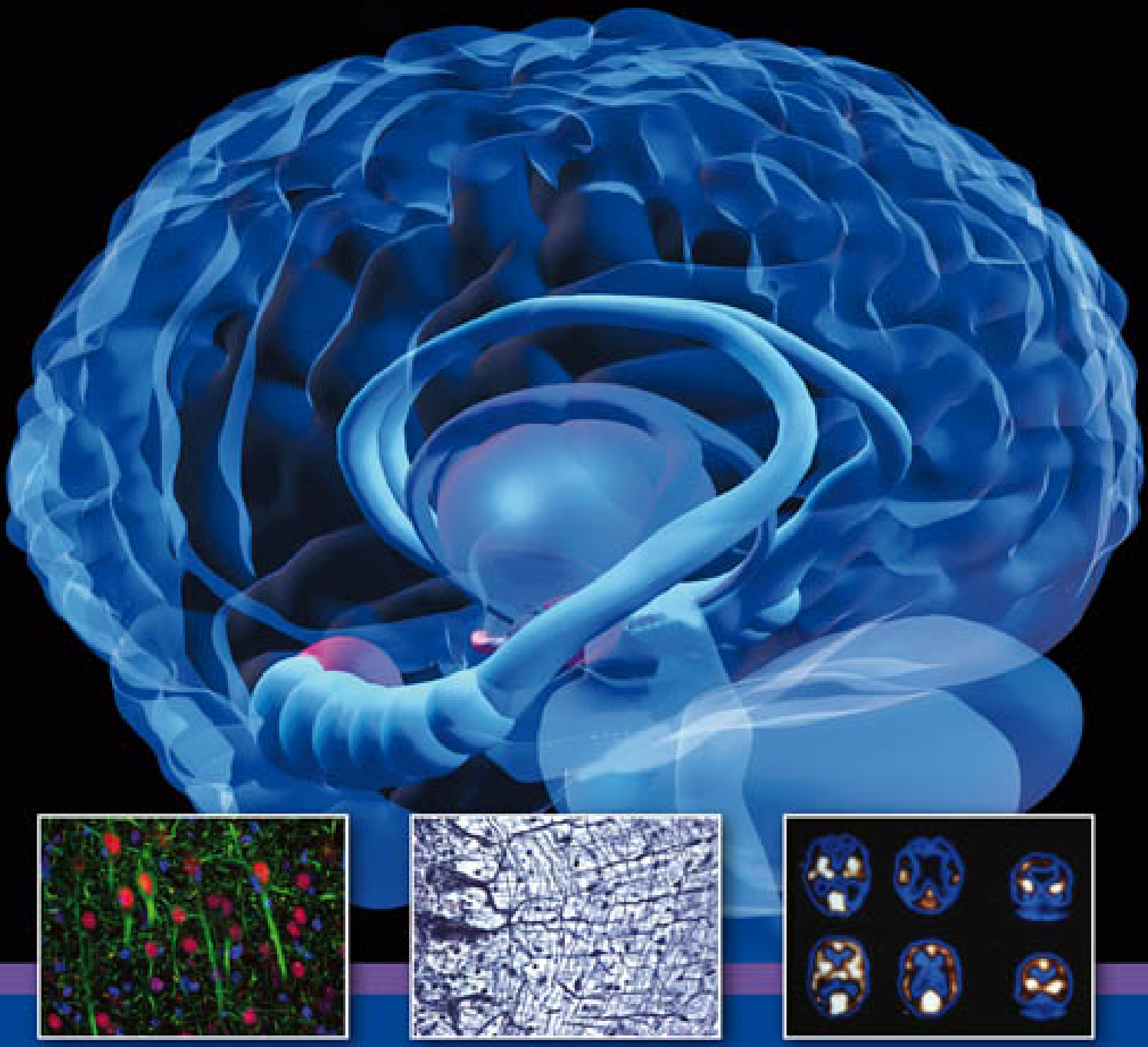

Psychiatry

An Evidence-based text

Basant K Puri and Ian Treasaden

Psychiatry:
An Evidence-Based Text

This page intentionally left blank

Psychiatry:
An Evidence-Based Text

Edited by

Professor Basant Puri
Honorary Consultant in Imaging and Psychiatry
Hammersmith Hospital and Imperial College, London, UK

Dr Ian Treasaden
Head of Forensic Neurosciences, Lipid Neuroscience Group
and Honorary Senior Lecturer in Psychiatry,
Imperial College London and Consultant Forensic Psychiatrist,
West London Mental Health NHS Trust, London, UK

HODDER ARNOLD
AN HACHETTE UK COMPANY

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20130319

International Standard Book Number-13: 978-1-4441-1326-6 (eBook - PDF)

Contents

Contributors

Angela Attwood BSc PhD
Department of Experimental Psychology, University of Bristol, Bristol, UK

James Barrett MB BS(Hons) BSc MSc FRCPsych
Lead Clinician, National Gender Identity Clinic, London, UK

RH Belmaker MD
Hoffer-Vickar Professor of Psychiatry, Ben Gurion University of the Negev, Beer-Sheva, Israel

Boben Benjamin BSc (Hons) MD LRCP LRCS LMSSA MRCPsych
Specialist Registrar in Old Age Psychiatry, Wythenshawe Hospital, Manchester, UK

Dinesh Bhugra MA MSc MBBS FRCPsych MPhil Phd
Professor of Mental Health and Cultural Diversity, Institute of Psychiatry, Kings College London, London, UK

Jonathan Bird BSc MBChB FRCPsych
Consultant Neuropsychiatrist and Clinical Electro-encephalographer, The Burden Institute, Frenchay Hospital, Bristol, UK

Roger Bloor MD MPsyMed FRCPsych
Senior Lecturer and Consultant Psychiatrist, Keele University Medical School, Harplands Hospital, Staffordshire, UK

Mary Boulton BA PhD HonMFPH
Professor of Health Sociology, School of Health and Social Care, Oxford Brookes University, Oxford, UK

Andrew Brittlebank MA FRCPsych
Director of Medical Education, Northumberland Tyne and Wear NHS Trust, St Nicholas Hospital, Newcastle upon Tyne, UK

Ian Brockington MB B Chir MPhil MD FRCP FRCPsych
Professor Emeritus, University of Birmingham

Caroline Brown BSc PhD
Senior Lecturer, Department of Psychology, University of the West of England , Bristol, UK

Peter F Buckley MD
Professor and Chairman, Department of Psychiatry, Associate Dean for Leadership Development, Medical College of Georgia, Augusta, Georgia, USA

Alistair Burns MBChB FRCP FRCPsych MD MPhil DHMSA
Professor of Old Age Psychiatry, Deputy Dean (Clinical Affairs) University of Manchester, Manchester, UK

Stuart Carney MBChB MPH MRCPsych FAcadMed
Deputy National Director, UK Foundation Programme, Cardiff, UK

Charlotte Clark BSc (Hons) PhD
Lecturer in Environmental and Mental Health Epidemiology, Centre for Psychiatry, Wolfson Institute of Preventive Medicine, Barts and the London School of Medicine & Dentistry, Queen Mary, University of London, London, UK

Ilana Crome MA MD MPhil FRCPsych
Academic Director of Psychiatry and Professor of Addiction Psychiatry, Keele University Medical School, Harplands Hospital, Staffordshire, UK

Sean Cross BSc (Hons) MBChB MSt MRCPsych
SpR, Department of Liaison Psychiatry, St Thomas' Hospital, London, UK

Rudi Dallos BA MSc PhD Dip (Clin Psych)
Department of Clinical Psychology, University of Plymouth, UK

Mark Dickinson BSc MBBS MRCPsych
The Combined Learning Disability Team, Enfield, Middlesex, UK

Timothy G Dinan MD PhD FRCPsych
Department of Psychiatry, University College Cork, Ireland

Kedar Nath Dwivedi MBBS MD DPM FRCPsych
Director, International Institute of Child and Adolescent Mental Heath, Visiting Professor, London Metropolitan University, Formerly, Consultant Child and Adolescent Psychiatrist, Northampton, UK

Nicol Ferrier BSc MD FRCPsych FRCPed
Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK

Peter Fitzgerald MB MRCPsych
Department of Psychiatry, University College Cork, Ireland

Peter Fonagy PhD FBA
Freud Memorial Professor of Psychoanalysis, University College London, Chief Executive, The Anna Freud Centre, London, UK

Bill (KWM) Fulford DPhil FRCP FRCPsych
Fellow of St Cross College, Member of the Philosophy Faculty and Honorary Consultant Psychiatrist, University of Oxford; Professor of Philosophy and Mental Health, University of Warwick; Co-Director, Institute for Philosophy, Diversity and Mental Health, University of Central Lancashire; and Special Adviser for Values-Based Practice, Department of Health, London, UK

Peter Garrard PhD FRCP
Reader in Neurology, University of Southampton School of Medicine Division of Clinical Neuroscience Southampton, UK

Steve Gentleman BSc PhD
Deputy Head of Division (Teaching), Reader in Experimental Neuropathology, Department of Clinical Neuroscience, Division of Neuroscience & Mental Health, Imperial College London, London, UK

Michael Gill MD MRCPsych FTCD
Professor of Psychiatry and Head of Department, Department of Psychiatry, Trinity Centre for Health Sciences, St. James' Hospital, Dublin, Ireland

Sandy Goldbeck-Wood MBChB MIPM MA (Psychoanalytic Studies) Dip (ERT)
Specialist Doctor in Obstetrics and Gynaecology, Ipswich Hospital, Suffolk, UK

John Gordon AB (Harvard)
Consultant Adult Psychotherapist, Forensic Psychotherapy Department and the Cassel Hospital, West London Mental Health NHS Trust; Psychoanalyst (British Psychoanalytic Association); Group Analyst (Institute of Group Analysis); Lecturer, Faculty of Continuing Education, Birkbeck College, University of London; and Honorary Senior Lecturer, Imperial College Medical School, London, UK

John Green BA MScClinPsych PhD
Chief Psychologist, CNW Department of Clinical Health Psychology, St. Mary's Hospital, London NHS Foundation Trust, UK; and Honorary Senior Lecturer in Behavioural Sciences, Imperial College, London

Richard Gross BA (Jt Hons) MA (Ed) Cert Ed
Full time academic author of many widely-used Psychology texts

Michael K Harte BSc PhD
Division of Pharmacology, School of Pharmacy, University of Bradford, Bradford, UK

Chris Hawley MB BS MRCPsych
Hertfordshire Partnership Foundation Trust, St Albans; and Bedfordshire and Hertfordshire Postgraduate Medical School, University of Hertfordshire, Hatfield, UK

David Hewison LLB PGDipSW CQSW MA ProfMbr SAP DipMarPsych FMbr SCPP DCpIPsychPsych CertMgmt (Open)
Senior Couple Psychoanalytic Psychotherapist and Head of Research, Tavistock Centre for Couple Relationships, London, UK

Frank Holloway FRCPsych
Consultant Psychiatrist and Clinical Director, Croydon Integrated Adult Mental Health Service, South London and Maudsley NHS Foundation Trust, Bethlem Royal Hospital, Beckenham, UK

Malcolm Hooper PhD BPharm MRIC CChem
Emeritus Professor of Medicinal Chemistry, University of Sunderland, Sunderland, UK

Byron Hyde MD
Chairman, Nightingale Research Foundation, Ottawa, Canada

Borislav Iankov MBBS MRCPsych
Specialty Training Registrar – General Adult Psychiatry, Bedfordshire & Luton Mental Health and Social Care Partnership Trust, Luton, UK

Trisha A Jenkins BSc PhD
Department of Psychiatry, Queen's University Belfast, Belfast, UK

Peter B Jones MD PhD FRCIP FRCPsych FMedSci
Professor of Psychiatry and Head of Department, Department of Psychiatry, University of Cambridge, Cambridge, UK

Sridevi Kalidindi MRCPsych
Consultant Psychiatrist in Recovery and Rehabilitation Psychiatry, Croydon Integrated Adult Mental Health Service, South London and Maudsley Foundation NHS Trust, Bethlem Royal Hospital, Beckenham, UK

Gabriel Kirtchuk
Consultant Psychiatrist in Psychotherapy, Psychotherapy Training Department, West London Mental Health NHS Trust, Southall, Middlesex, UK

J Hubert Lacey MD MPhil FRCPsych
Professor of Psychiatry, St George's, University of London and Director of St George's Eating Disorders Service, London

Malcolm Lader OBE LLB DSc MD PhD FRCPsych FMed Sci
Emeritus Professor of Clinical Psychopharmacology, Institute of Psychiatry, King's College London, UK

Morven Leese MSc PhD
Health Service and Population Research Department, Institute of Psychiatry, Kings College London, London, UK

Pesach Lichtenberg MD
Department of Psychiatry, Herzog Hospital; and Clinical Senior Lecturer in Psychiatry, Hadassah Medical School of the Hebrew University of Jerusalem, Israel

Geoffrey G Lloyd FCC Psych MD MA
Visiting Consultant Psychiatrist, Priory Hospital North London, London, UK

Brian Lunn MB ChB FRCPsych
Honorary Clinical Senior Lecturer School of Medical Sciences Education Development, Newcastle University, Newcastle upon Tyne, UK

Chris Mace MD FRCPsych
Consultant Psychotherapist, Coventry and Warwickshire Partnership Trust, Honorary Associate Professor in Psychotherapy, University of Warwick, UK

Hameen Markar FRCP FRCPsych MPhil
Medical Director, Bedfordshire & Luton Mental Health and Social Care Partnership NHS Trust, Luton, UK

Rebecca Martinez DM MSc MRCEP MRCPsych
Clinical Lecturer in Psychiatry, Faculty of Medicine, University of Glasgow, Glasgow, UK

Jane McGrath MB MRCPsych
Department of Psychiatry, Trinity Centre for Health Sciences, St. James' Hospital, Dublin, Ireland

C Susan Mizen MBBS MRCPsych Memb SAP
Consultant Psychiatrist in Psychotherapy, Wonford House Hospital, Exeter, UK

David P Moore MD
Associate Clinical Professor, Department of Psychiatry and Department of Neurosurgery, University of Louisville School of Medicine, Louisville, Kentucky, USA

Ann Mortimer BSc MMedSc MD FRCPsych
Foundation chair in Psychiatry and Head of Department of Psychiatry, University of Hull, Hull, UK

Victoria Mountford BA Hons DClinPsych
Clinical Psychologist and Honorary Research Fellow, St George's Eating Disorders Service and St George's, University of London

Marcus R Munafò PhD
Reader in Biological Psychology, Department of Experimental Psychology, University of Bristol, Bristol, UK

Kingsley Norton MA (Cantab) MD FRCPsych
Consultant Psychotherapist, West London Mental Health Services, Southall, Middlesex, UK

Ash Padhi SpR
Specialist Trainee, Hertfordshire Partnership Foundation Trust, St Albans, UK

Hilary Pavis MA MRCGP
Consultant in Palliative Medicine, Nightingale Macmillan Unit, Royal Derby Hospital, Derby, UK

Raj Persaud FRCPsych
Consultant Psychiatrist, Emeritus Visiting Gresham Professor for the Public Understanding of Psychiatry, Gresham Collge, London, UK

Gavin P Reynolds BA PhD
Professor of Neuroscience, Department of Psychiatry, Queen's University Belfast, Belfast, UK

Allan Scott BSc MPhil MD MBA FRCPsych
Consultant Psychiatrist, Royal Edinburgh Hospital, Edinburgh, UK

John Shneerson MA DM MD FRCP FCCP
Sleep Disorders Unit, Papworth Hospital, Cambridge, UK

Iqbal Singh MB BS FRCPsych
Consultant Psychiatrist in Learning Disability, The Riverside Centre, Hillingdon Hospital, Uxbridge, Middlesex, UK

Georgina Smith
Department of Clinical Health Psychology, St. Mary's Hospital, London, UK

Paul St John Smith MA (Oxon) BM BCh FRCPsych
Consultant Psychiatrist, Hertfordshire Partnership Foundation Trust, St Albans, UK

Daniel Stahl Dipl Biol Dr rer nat
Lecturer in Biostatistics, Dept of Biostatistics & Computing, Institute of Psychiatry, Kings College London, London, UK

Stephen Stansfeld MB BS PhD MRCP FRCPsych
Chair & Centre Lead for Psychiatry, Centre for Psychiatry, Wolfson Institute of Preventive Medicine, Barts and the London School of Medicine & Dentistry, Queen Mary, University of London, London, UK

Maelie Swanwick FRCP
Consultant in Palliative Medicine, Nightingale Macmillan Unit, Royal Derby Hospital, Derby, UK

David Taylor BSc FFPH
Chief Pharmacist, South London and Maudsley NHS Foundation Trust, London, UK

Tim Thornton
Director of Philosophy and Professor of Philosophy and Mental Health, International School for Communities, Rights and Inclusion, University of Central Lancashire, Preston, UK

Sami Timimi MBChB MRCPsych
Consultant child and Adolescent Psychiatrist, Lincolnshire Partnership NHS Trust, Visiting Professor of Child and Adolescent Psychiatry, Faculty of Health and Social Sciences, Lincoln University

Lorna Torrens BA DClinPsych
Head of Clinical Psychology, Astley Ainslie Hospital, Edinburgh, UK

Trevor Turner DHMSA MD FRCPsych
Consultant Psychiatrist, East London Foundation Trust, Homerton University Hospital London; and Honorary Senior Clinical Lecturer, Barts and the London School of Medicine & Dentistry, Queen Mary, University of London, UK

Peter Tyrer FMedSci FRCPsych
Department of Psychological Medicine, Imperial College London, London, UK

Stephen Tyrer MB LMCC FRCPsych
Consultant in Psychiatry and Pain, Mental Health Unit, Southland Hospital, Invercargill, New Zealand

Fiona Warren MA PhD
Department of Psychology, Faculty of Arts and Human Sciences, University of Surrey, Guildford, UK

Kate Webb MRCPsych
Sp R and Honorary Research Assistant, St George's Eating Disorders Service and St George's, University of London, London, UK

Christopher J Williams BSc MBChB MMedSc MD FRCPsych
Senior Lecturer in Psychiatry, Faculty of Medicine, University of Glasgow, Glasgow, UK

Ewa Wisniewska Young MRCPsych
Specialist Registrar, Affective Disorders Unit, Royal Victoria Infirmary, Newcastle upon Tyne, UK

Ewa Zadeh MRCPsych
SpR and Honorary Research Assistant, St George's Eating Disorders Service and St George's, University of London, UK

Adam Zeman DM FRCP
Professor of Cognitive and Behavioural Neurology, Peninsula Medical School, Exeter, UK

Preface

Among the many demands on modern psychiatrists is keeping up to date with the ever-increasing pace of the development of psychiatric knowledge in recent years, both in the biological and psychosocial spheres. The basic sciences of psychology, neuroanatomy, neurophysiology, neuroendocrinology, neuroimaging, neuropathology, genetics, biochemistry, pharmacology, neuroscience, epidemiology and social sciences have now led to a better understanding of the basic mechanisms underlying clinical disorders. Sub-specialties of psychiatry have rapidly developed and the practice of psychiatry has increasingly moved from institutions to the community. Clinicians are now under pressure to deliver high quality cost-effective patient-focused care based upon the best evidence available.

The aim of this volume is to provide an up-to-date, solid, evidence-based text. Many of our contributors are acknowledged international leaders in their respective fields, have often been centrally involved at the forefront of shaping psychiatric research and practice and have, as a result, a first-hand feel of the evidence base of their contributions. The strength of the evidence base does still vary widely in different fields of psychiatry, but with the development of national guidelines, systematic reviews and meta-analyses, individual, sometimes idiosyncratic, clinical practices of the past are now no longer acceptable unless based on evidence. Thus, even if not entirely evidence based, psychiatry should always be evidence informed. Psychiatrists who increasingly work in multidisciplinary teams and whose practice is now increasingly challenged by other professionals, managers and, indeed, patients, now have to be able to defend the evidence base to their practice, if they are to maintain their medical leadership role in psychiatric practice.

In this book, we aim to comprehensively describe the basic sciences and clinical disorders and their treatments – using the UK and Ireland MRCPsych syllabus as a guide – while emphasising the evidence underlying theory and practice for the topics covered. However, we believe that this book will cater not only for trainee psychiatrists studying for the MRCPsych examinations, but also for trainees elsewhere and, indeed, will provide a valuable resource for psychiatrists who have completed their training and other professionals who work in psychiatry.

To facilitate the aim of this project, the book is divided into major sections and 79 chapters in a carefully considered order. Chapters have been standardised and cross referenced and include important and up to date references and generous use of tables, figures, boxes and pictures. At the end of each chapter major learning points are identified. While the book strives to provide an integrated overview of current knowledge through its sections and chapters, ,chapters have also been designed to stand alone, which inevitably implies some overlap in content between chapters which we hope has been kept to an acceptable minimum.

We hope this book will achieve wide acceptance through its succinct, user-friendly approach and its recognition of the importance of a solid evidence base for psychiatric practice. While a text book alone does not make a good psychiatrist, we hope this one will provide the sound foundation of evidence-based theoretical knowledge required for the competent practising clinician of today.

Basant K Puri
Ian Treasaden

Acknowledgements

We would like to warmly thank our contributors for their excellent contributions, delivered often on tight deadlines. We would also like to thank our publishers Hodder, including the Commissioning Editor, Philip Shaw, and the Project Editors, Amy Mulick and Joanna Silman, for their encouragement and always helpful assistance. We also wish to thank our Publisher, Caroline Makepeace. We would additionally like to thank Susan Oxlade for her administrative and secretarial assistance.

PART 1

The foundations of modern psychiatric practice

This page intentionally left blank

Chapter 1

History of psychiatry

Trevor Turner

INTRODUCTION

Psychiatric illness seems intrinsic to – and a necessary part of – being *Homo sapiens*, so it is not surprising that there are descriptions of abnormal behaviour, and even of symptoms resembling schizophrenia, in the earliest written fragments of Egypt, India and Babylon. Some evolutionary theorists consider psychosis to be inextricably linked to the genetic mutation(s) for our enlarged frontal lobe, with its unique capacity for memory and imagination. A history of psychiatry has to understand therefore how 'mental' symptoms – presentations of madness, sadness, foolishness, strangeness – were understood in all societies and accepted, or not, and what was done, or not, to help the sufferers.

There is thus a long history of mental illness consisting of 3000 years of random descriptions (in religious writings, legal documents, diaries, histories, romances and plays), and a shorter and more technical/medical history of practical psychiatry, in terms of professionalized attempts to understand, diagnose and treat mentally ill people. The long history precedes even the Bible and Classical Greece, while the short history starts some 200 years ago (at least in European terms), a product of the scientific advances of the Enlightenment, industrialization and city life. Supernatural 'apparitions' become secularized 'hallucinations', and psychopathology begins to erode the varieties of religious explanation.

HISTORIOGRAPHY – i.e. WHAT'S BEEN WRITTEN ABOUT IT

This chapter concentrates on the past two centuries as being essential to understanding what drives modern psychiatry. Just as taking a history is the key to psychiatric diagnosis, so knowing how and why we are at our present state of knowledge and organization helps us to make sense of our current work. In itself, the topic 'the history of psychiatry' has undergone a radical change in the past 30 years, from the obscure jottings of retired psychiatrists to a thriving subspecialty of modern historical research. It has its own quarterly journal, *History of Psychiatry* (first published in March 1990), its own European Association and a body of dedicated academic historians. The standard heavyweight tomes, such as *The History of Psychiatry* by Alexander and Selesnick (1966), have been replaced by more svelte versions, such as Edward Shorter's (1997) *A History of Psychiatry: From the Era of the Asylum to the Age of Prozac*, and diagnosis-based compilations such as *A History of Clinical Psychiatry: The Origin and History of Psychiatric Disorders* (Berrios and Porter, 1995). There has also been a series of reprints of core psychiatric textbooks and careful descriptions of whole eras of change, for example Andrew Scull's (1979) *Museums of Madness*, subtitled *The Social Organisation of Insanity in Nineteenth Century England*. Details of the most substantial histories of psychiatry, and the most influential secondary texts and textbooks (or versions thereof) are given in Tables 1.1–3.

Most importantly, there has been a prolonged, sometimes bitter but generally creative debate since the rise of the so-called 'anti-psychiatry movement' of the 1960s and 1970s as to what is even meant by a 'history of psychiatry' or a 'history of insanity' or of 'madness', and so on. Traditionally it had been the recording of notable forms of mental illness, whether in literature (e.g. Shakespeare's King Lear or Othello) or medical writings, accompanied by outlines of new diagnostic or treatment approaches, with admiring biographies of the leading doctors (not always medical) who had tried to help or even 'cure' the afflicted mad folk. This so-called 'meliorist' history – things getting better, in terms of more accurate diagnoses, more thoughtful doctors (and attendants/nurses) and more humane treatments – was challenged in the writings of social theorists such as Michel Foucault and Andrew Scull (see Table 1.2).

Foucault postulated a 'Great Confinement' in the seventeenth and eighteenth centuries, whereby the world of free-thinking and imaginative 'unreason' had been corralled by the mechanistic warriors of reason and social control. Psychiatrists and the psychiatric enterprise were even seen as a kind of thought police of this dark Enterprise. Mental illness was theorized as 'socially constructed', a means of rounding up the deviants and low life in society rather than a genuine part of medical practice dealing with real diseases. This debate goes on, has involved historians, psychologists, anthropologists, lawyers, sociologists and even a few psychiatrists, and has now swung back to some degree

Table 1.1 Ten histories of psychiatry

Tuke DH (1968; reprint of 1882 edn) *Chapters in the History of the Insane*, Amsterdam: Bonset	The original heroic nineteenth-century history, written as a celebration of Victorian achievement in building asylums and rescuing 'lunatics' from the neglect and abuse of whips, chains and supernatural beliefs
Zilboorg G, with Henry GW (1941) *A History of Medical Psychology*, New York: Norton & Co.	Zealously pro-Freudian ('the second psychiatric revolution'), this is written in the classic 'great men' style of history but has good details and nice photographs. Psychoanalysis is seen as the cure for all mental illness and more
Hunter R, Macalpine I (1963) *Three Hundred Years of Psychiatry 1535–1860*, Oxford: Oxford University Press	An extraordinary collection, over 1000 pages long, of extracts from several hundred original (primary) sources arranged chronologically and with a superb index. The ideal reference book for any facet of psychiatric diagnosis, language and treatment. Great for lecture introductions
Alexander FG, Selesnick ST (1966) *The History of Psychiatry*, New York: Harper & Row	The standard of its time but also in awe of Freud, more than half being devoted to the 'Freudian Age'. Much material from classical/medieval times, but unreliable in details. For example: 'It does not seem probable that ... pharmacological and biological methods will throw more light upon the complex phenomena of interpersonal relationships and replace psychological methods of treatment'
Jones K (1993) *Asylums and After: A Revised History of the Mental Health Services – From the Early Eighteenth Century to the 1990s*, London: Athlone Press	Describing how services developed, this is the best analysis of the rise and fall of the asylums, with special insight into the social, legal and political influences on psychiatry, including the deinstitutionalization of the past 50 years. Very critical of the weaknesses of 'care in the community'.
Scull A (1993) *The Most Solitary of Afflictions: Madness and Society in Britain 1700–1900*, New Haven, CT, and London: Yale University Press	An updated version of the author's groundbreaking *Museums of Madness* (1979), which introduced a radically new take on psychiatry as representing social power and social control, thus reinforcing the status quo via an often doubtful construct of 'mental illness'. A challenging way of looking at history, even if you disagree
Shorter E (1997) *A History of Psychiatry: From the Era of the Asylum to the Age of Prozac*, New York: John Wiley & Sons	The best modern history by far, though avowedly organicist in tone. Readable, beautifully detailed and with real narrative drive, concentrating on developments since *c.*1800. The author concludes that 'neurochem' and 'neurochat' augment each other as the optimum form of care
Andrews J, Briggs A, Porter R, Tucker P, Waddington K (1997) *The History of Bethlem*, London: Routledge	A highly detailed and extensively researched 750 pages of modern social history, published to celebrate Bethlem Hospital's 750th anniversary. A very technical text, for the specialist
Berrios G, Porter R (eds) (1995) *A History of Clinical Psychiatry: The Origin and History of Psychiatric Disorders*, London: Athlone	A multi-author volume reviewing the clinical and social history of our understanding of the major mental disorders (e.g. dementia, schizophrenia, anxiety), including neuropsychiatric conditions and the epilepsies. Not for reading through, but it is well referenced, and individual topic chapters (e.g. post-traumatic shock disorder, shell shock) are fascinating
Fuller Torrey E, Miller J (2001) *The Invisible Plague: The Rise of Mental Illness from 1750 to the Present*, Piscataway, NJ, and London: Rutgers University Press	A spirited and informative outline concentrating on the theory that severe schizophreniform mental illness represents a modern epidemic due to industrialization and urbanization. Diet, toxins, infectious agents (related to vaccination?) and even pet cats are among the postulated aetiological agents. Or is this retrospective history gone too far?

towards a general acceptance that there, probably, is such a thing as mental illness.

The key research has been into archival sources such as asylum records (going back to the mid-nineteenth century and sometimes earlier), the casebooks of individual doctors, for example the famed eighteenth-century 'mad-doctor' and physician to Bethlem ('Bedlam') Hospital, John Monro (1715–91), the writings of psychiatric patients themselves,

Table 1.2 Ten influential secondary texts

Foucault M (1965) *Madness and Civilization: A History of Insanity in the Age of Reason* (trans. R Howard), New York: Random House	The key work of the most celebrated cultural historian of the past 50 years, this posits a belief that mental illness is 'not a natural fact but a cultural construct', not a disease needing treatment but an issue of freedom, knowledge and power. Written in the impressionistic, deliberately indefinite style of much modern French philosophy, it is a real challenge, both to psychiatry as a reasonable profession and to the average empirical English reader
Hunter R, Macalpine I (1969) *George III and the Mad Business*, London: Penguin	Classic outline of good archival research into the much debated illness of King George III (1760–1820). Was it porphyria or just another routine manic disorder? More recently turned into a popular play and then a film, *The Madness of King George*, by the writer Alan Bennett
Ellenberger HF (1970) *The Discovery of the Unconscious: The History and Evolution of Dynamic Psychiatry*, New York: Basic Books	Over 900 pages on the development of psychological approaches to mental illness, and how Freud emerged successfully from numerous rivals. An absorbing matchless account of how theories come and go
Hunter R, Macalpine I (1974) *Psychiatry for the Poor: 1851 Colney Hatch Asylum–Friern Hospital 1973 – A Medical and Social History*, Folkestone: Dawson	Minutes, reports, photos, patients, medical and nursing staff – the archetypal asylum history, unvarnished but perfectly evocative of the practical management of an enormous institution (up to 2700 inmates) in a bygone era
Simon B (1978) *Mind and Madness in Ancient Greece: The Classical Roots of Modern Psychiatry*, Ithaca, NY: Cornell University Press	The classic exposition of Greek ideas, with chapters on 'tragedy and therapy' and 'Plato and Freud'
MacDonald M (1981) *Mystical Bedlam: Madness, Anxiety and Healing in Seventeenth Century England*, Cambridge: Cambridge University Press	A detailed analysis of the records of Richard Napier (1559–1634), a mystic and astrologer, much famed in his time, who treated over 2000 mentally disturbed patients who might be 'mopish', 'troubled in mind' or 'mad/lunatic'
Petersen D (ed.) (1982) *A Mad People's History of Madness*, Pittsburgh, PA: University of Pittsburgh Press	A collection of 27 accounts of being mentally ill, written by the patients, from the fifteenth to the twentieth century. A selective selection, but well edited historically. History from the patient's viewpoint, and so a unique version of common symptoms
Valenstein ES (1986) *Great and Desperate Cures: The Rise and Decline of Psychosurgery and Other Radical Treatments for Mental Illness*, New York: Basic Books	It will 'chill the marrow of naive readers' ran the blurb, and this is a well-researched account of lobotomy, its leading exponents (Moniz and Freeman) and its rationale. Imagine having your brain cut into as part of an out-patient procedure, and then being sent home to convalesce
Berrios GE, Freeman H (eds) (1991) *150 Years of British Psychiatry 1841–1991* London: Gaskell; and Berrios GE, Freeman H (eds) (1996) *Volume II: The Aftermath*, London: Athlone	A vivid collection of 28 essays, mainly by psychiatrists, covering the 'institutions', 'ideas' and 'people' dominating UK psychiatric development. Wonderfully, obsessionally(?) detailed. Much praised by professional historians
Porter R (2002) *Madness: A Brief History*, Oxford: Oxford University Press	A mere 230 or so pages. A brief, lively and easy-to-read historian's account of 'madness' – so not in fact confined to 'psychiatry'

and court records dealing with, for example, insane murderers. If you are prepared to hunt around in medieval documentation – and that is hard work, in terms of finding and even reading the material, which is often in Latin – and in dusty library records, a much more nuanced picture of how madness has been dealt with through the ages can be extracted, albeit often in bits and pieces, like the shards of pottery from an archaeological dig.

For example, if we consider how people deemed to be 'mad' or 'lunatic' were dealt with in Shakespeare's time (1564–1616), whether in his plays or in the plays of his contemporaries, it is not a lot different from today. Recognizing

Table 1.3 Ten interesting textbooks

Pinel P (1801) *A Treatise on Insanity* (trans. DD Davis), New York: Hafner Publishing	An outline of 'maniacal disorders', including an attempt at classification and numerous case histories. A humane, detailed and easy-to-read masterpiece – the first modern textbook with delightful descriptions of treatments, theatrical(s), and moral and physical approaches to managing very disordered people
Haslam J (1810) *Illustrations of Madness: Exhibiting a Singular Case of Insanity and a No Less Remarkable Difference in Medical Opinion*, London: Hayden; reprinted, Porter R (ed.) (1988) London: Tavistock Classics	A book-length account of a contended case, illustrating a 'first-rank' series of colourful symptoms typical of a florid paranoid schizophrenia, but that nevertheless generated considerable doubt and debate at that time
Bucknill J, Tuke DH (1858) *A Manual of Psychological Medicine*; facsimile edn (1968) New York: Hafner	The first proper English treatise of psychiatry, indicating the growing size of the specialty and the need for a student's textbook. Easy to read, quite sceptical, and with excellent clinical descriptions and references to the literature. Treatment is divided into 'hygienic', 'moral' and 'medical'
Maudsley H (1867) *The Physiology and Pathology of the Mind* London: Macmillan	A much admired textbook outlining the physical basis of mental disease as opposed to the 'metaphysical' theorizing that tended to dominate public discussion. Maudsley was a bluff Yorkshireman with a sarcastic tongue, but he was widely read in the French and German literature. Although much quoted, it was probably not practically used by working psychiatrists of his time
Von Krafft-Ebbing R (1886) *Psychopathia Sexualis: With Especial Reference to Contrary Sexual Instinct – A Medico-Legal Study* (trans. G Chaddock 1892), Philadelphia, PA: Davis & Co.	The first detailed description of abnormal sexual behaviours, including sadism, masochism, 'congenital inversion' (i.e. homosexuality) and fetishism. An entertaining collection of case histories, in great detail, bringing to light a wide range of sexual 'pathologies' usually suppressed even in the technical and medical literature of the time
Freud S (1900) *The Interpretation of Dreams* (trans. J Strachey 1976), London: Pelican Books	The classic Freud text on his theory of the unconscious, dreams being considered essential to understanding one's inner mental life. From this he moved on to developing his highly influential theories of childhood sexuality and the 'Oedipal complex' of character development. Lots of examples, but not 'scientific', although claimed as such
Kraepelin E (1904, 1905, 1912 – very popular) *Lectures on Clinical Psychiatry* (ed. T Johnston), London: Bailliere, Tindall and Cox	Summarizing the classic and groundbreaking textbook – Kraepelin in fact produced eight editions of his Lehrbuch from 1886 to 1924, which first described dementia praecox as a distinct psychotic disorder. His distinction between dementia praecox/schizophrenia and manic-depressive insanity continues to dominate diagnostic categories today
Bleuler E (1911) *Dementia Praecox or The Group of Schizophrenias* (trans. J Zinkin 1950), New York: International University Press	The book that introduced the term 'schizophrenia', fusing psychoanalytic theory derived from Freud with the clinical descriptions of Kraepelin. The Bleulerian outline of schizophrenia dominated psychiatry until the 1960s. His four 'A's (autism, affective impairment, ambivalence, impaired associations) made reliable diagnosis rather difficult
Sargant W, Slater E (1944) *An Introduction to Physical Methods of Treatment in Psychiatry*, Edinburgh: Livingstone	The (wartime) classic of biological psychiatry, trumpeting the use of insulin therapy, electroconvulsive therapy (ECT), chemical sedation, malaria treatment and prefrontal leucotomy, as opposed to psychotherapy, to which the authors barely paid lip service. Biology red in tooth and claw, claiming mental disorders as 'dependent on physiological changes'
Jung CG (1963) *Memories, Dreams, Reflections*; reprinted (1995) London: Fontana Press	The summation of the theories and work of Carl Jung (1875–1961), the long-lived alternative to Freud. Many 'Jungian analysts' are still influential today. His notion of archetypes and the collective unconscious are at the very edge of psychiatry, verging into spirituality, anthropology and religion. And he did actually try to psychoanalyse psychotics, unlike Freud, as outlined in his 1907 *The Psychology of Dementia Praecox*

mental illness was based on people's behaviours, what they said and what they perceived (e.g. Lady Macbeth's recurrent and obsessional hand-washing), doctors are called to attend them (e.g. Dr Caius in *The Comedy of Errors*), and disordered brains are seen as part of the problem ('canst thou not minister to a mind diseased … raze out the written troubles of the brain …', asks Macbeth of his physician). Confinement in madhouses, or 'dark houses' (e.g. Malvolio in *As You Like It*), or somewhere out of the way, is a common resort. We even have, in the character of Tom O'Bedlam in the play *King Lear* a vivid description of someone pretending to be mad. This is the eventual hero, Edgar, who in order to avoid entrapment and murder hides out half-naked, gibbering to himself and expressing a range of symptoms that today are entirely consistent with what we call 'schizophrenia'. If that is what early-seventeenth-century play-goers recognized as mental illness, then it seems reasonably indicative of the persistence of what we call schizophrenia, over some 400 years or more, rather than it being (as some have postulated) a relatively new disease of the industrial age of the nineteenth century and beyond (see especially Fuller Torrey, in Table 1.1).

TREATING MENTAL ILLNESS: WHERE, HOW, WHO?

The past 200 years have seen the rise and fall of a standard approach to mental illness – the asylum. Founded as a vehicle for delivery of what was essentially a structured psychosocial intervention, namely the 'moral treatment', the asylum has dominated psychiatric practice since the mid-nineteenth century – throughout the world – and only in the past 30 years has it became subservient to the notions of 'community care'. In fact the asylum arose because of the failures of community care – troubled souls wandering the countryside and dying of starvation, or being chained up in cellars or attics (à la Rochester's wife in the 1847 novel *Jane Eyre*) or unreliable private madhouses. The combination of Quaker concern, evangelical Christianity and the bricks-and-mortar solution of Victorian industrialization demanded that the 'pauper insane' be placed in large sanctuaries, i.e. the asylums, rather like medieval lepers in leprosaria. The rise of modern risk management and the medium-secure unit (MSU) are merely a replication of this process under a new guise. 'Normal' folk distrust (and exploit) 'mad' folk – thus the unique role of the psychiatrist in negotiating the relationship.

There is also during this period the development of a specialist profession, from a few scattered individuals, good-hearted clergymen looking after people in their houses, to the organization of medico-psychological associations and the first regular journals (e.g. *The Asylum Journal* in 1854, now the *British Journal of Psychiatry*). The troubled story of the development of psychiatry reflects another essential theme, namely stigma, the fear of the insane, and the difficulties their case has met in terms of resources, trained staff, research-based treatments and integration back into 'normal' society. Even today many doctors and nurses qualify, all over the world, without having partaken in any genuine psychiatric training or practice.

Yet psychiatry, one of the first true specializations in medicine, actually spawned the subspecialty of neurology and is the branch of medicine most outward-looking in terms of its relationship with sociology, psychology, criminology, the law, architecture, and social and working life. It has its heroes and its villains of course, more of the former than the latter, and has been cruelly mocked, not least in modern cinema (e.g. Hannibal Lecter in *The Silence of the Lambs* in 1991) and for indulging in what are termed 'desperate remedies' such as lobotomy and electroconvulsive therapy (ECT). But much retrospective criticism, which rarely comes from mainstream historians, often fails to understand how difficult it always has been to cope with chronic, unremitting, progressive psychosis.

The journey from essentially social and institutional treatments (e.g. asylums) to various psychological approaches (e.g. Sigmund Freud and friends), and on to the 1950s rise of psychopharmacology, is full of iconic stories. Philippe Pinel (1745–1826), the father of French psychiatry, unchaining the insane in 1792, in the middle of the chaos of the French Revolution (a battalion of soldiers hiding round the back of the hospital in case all hell broke loose); John Conolly (1794–1866) introducing, against mocking scepticism, non-restraint to the enormous Hanwell asylum (his monograph on non-restraint was published in 1856); the unravelling of 'general paralysis of the insane' (GPI) as in fact an infectious disease, tertiary syphilis, and its response to malaria therapy researched and developed by Julius Wagner-Jauregg (1857–1940), psychiatry's first Nobel Prize winner; the shell shock debate during and after the First World War (was it physical, a form of brain damage, or nervous, or cowardice even?), and the million or more war pensions generated; the transformed acute wards of the 1950s and the reported great interest of the nursing personnel when chlorpromazine, a 'neuroleptic', was introduced. These are mainly true, and some of the key historical themes are outlined in Table 1.4.

THE WORLD OF BEDLAM PRE-1800

Biblical and classical writings give many descriptions of people going mad, mainly kings and heroes such as Saul, Nebuchadnezzar and Ajax (who ran amok and slaughtered a herd of sheep). The long-influential Greco-Roman physician Galen (c.AD129–200) reinforced the Hippocratic doctrine of the humoral theory – the four humours (i.e. visible bodily secretions) being phlegm (tears or sweat), blood, yellow bile (choler) and black bile (= 'melancholia' in Greek), imbalances of which were seen as the basis of all

Table 1.4 Key historical themes

Language	From 'mania' and 'melancholia' to 'madness', 'insanity' and 'mental illness' – via 'neurosis' and 'psychosis' – there has been a constant and often confusing expansion of terms. Does this represent a changing pattern of illness or simply changing presentations? (Are mental illnesses cultural constructs or real diseases?) James Prichard's descriptions of cases of 'moral insanity', in 1835, would be diagnosed today as either personality disorder or bipolar illness. George Beard's 1869 outline of 'neurasthenia' is now covered by various forms of depression, anxiety and phobias. Modern 'schizophrenia' is relatively unique as a diagnostic term, surviving for over 100 years, unlike most of its contemporaries
Causation	Theories of humoral imbalance (i.e. some form of physical disorder) vied with supernatural beliefs (curses, witchcraft, secret sins) well into the eighteenth century. By the nineteenth century concerns about 'moral' causes (i.e. bad habits, sexual excess, cigar-smoking) vied with increasingly physical theories, as indicated by the practice of phrenology, with leading exponents being Drs Franz Gall (1758–1828) and JC Spurzheim (1776–1832). Phrenology considered the brain as the organ of the mind, different activities being located in different areas (the precursor of cerebral localization), demanding careful examination therefore of the shape of one's head (e.g. the notion of high-brow *v.* low-brow). This led on to the criminal anthropology of Cesare Lombroso (1836–1909) identifying criminality (and mental illness) via various physical stigmata. The ongoing debates as to nature *v.* nurture and organic *v.* psychological reflect this long-standing argument. The journey of the 'psychoses' from being disorders of the soul (i.e. psyche) to established diseases of the brain nicely illustrates this history
Treatment	Psychiatrists have always been asked to talk to patients in a special way ('fixing with the eye') in order to persuade them out of their delusions or into sober personal habits, voluntarily or by overt or covert enforcement. They have used 'behavioural' medications (e.g. to make one sweat or vomit), physical controls such as straitjackets or special chairs, locked wards and controlled diet and exercise, and social activities or even amusements, in a community context. Specific antipsychotics or monitored psychological therapies, as in the last half-century, are genuinely innovative. Putting mentally ill people into prisons is a reversion of Enlightenment principles
Law	When James Hadfield shot at King George III in a theatre on 15 May 1800 and was charged with 'high treason', he stated that he had been acting on God's instructions. Deemed not responsible, he was sent to Bethlem, and deciding on whether someone is 'mad' or 'bad' has subsequently dominated public attitudes to mental illness. The attempted murders of royalty (Queen Victoria endured at least seven failed assassinations), and murders by and of famous people, have always made excellent journalist copy. The McNaughton Rules, determining criminal insanity (e.g. 'knowing the nature of the act'), derive from the famous 1843 trial of Daniel McNaughton who murdered the then prime minister's private secretary, mistaking him for the prime minister. The 1863 founding of Broadmoor Hospital, for the reception of criminal lunatics, institutionalized the acceptance that capital punishment was not appropriate for people labouring under any 'defect of reason'. Henry Maudsley's 1874 *Responsibility in Mental Disease* was the first forensic psychiatric textbook, and the 'insanity defence' has been a key arbiter of psychiatry's public standing
Nervous disorders	Patients and practitioners have constantly wanted to distance themselves from 'raving and furious' madmen, even though they might seek help in dealing with chronic anxiety, tiredness, 'nerves', 'vapours', etc. The rise of a middle class generated a need for physicians specializing in 'office practice' (as opposed to asylum management). Mesmerism and hypnotism, spas and 'the water treatment', tonics and electric machines, and special dietary regimes such as Silas Weir Mitchell's 'Fat and Blood', a disgusting diet consisting of just those products while bed-bound for 6 weeks, could be fashionable and lucrative. The success of Sigmund Freud and other psychotherapists reflected this search for special, personal understanding. The ultimate mystique of the doctor therapist can be seen in TS Eliot's 1950 play *The Cocktail Party*

disease, physical or mental. Most of the latter came under the terms 'mania' and 'melancholia', people being overexcited by too much blood/choler or slowed down by too much phlegm or black bile. In the West the rise of Christianity led to the elaboration of supernatural and later witchcraft theories, mental illness even being seen as a punishment for wickedness, despite Jesus Christ 'casting out devils' from madmen, as regularly reported in the four Gospels.

From the fall of Rome to the Renaissance, such notions became a fixed part of approaches to treatment. The agrarian poor looked to holy shrines, magic potions, wise women or necessary beatings. The 'furious' mad were tied up in cellars or driven from parish to parish. Suicide was a sin – with a loss of possessions – and thus coroners' juries increasingly tried to establish a *non-compos mentis* verdict, medicalizing the deed in order to avoid penalizing the family. Physicians might be sought for the wealthy (Henry

VI, king of England from 1422 to 1461, when he heard 'voices' saw a learned doctor), with the use of laxatives ('purging') or bleeding as the standard approaches (to get rid of the excess humours).

With the Renaissance came a questioning of accepted ideas, with witchcraft trials (some 50 000 executions during the sixteenth and seventeenth centuries) providing one battleground. This European 'witch-craze' was fuelled by challenges to papal authority, for example by 'Protestant' preachers such as Martin Luther (1483–1546) or by scientists, such as Galileo Galilei (1564–1642) demonstrating that the earth went round the sun (not the other way round). Reactionary work appeared, famously *Malleus Maleficarum* (1486) written by two monks, Kraemer and Sprengler. A kind of handbook for witch-finding, this 'Hammer of Evil-Doing' was like a modern National Institute for Clinical Excellence (NICE) guideline, and it took a brave German physician, Johannes Weyer (1515–88), to produce the sceptical counterblast. His *De Praestigiis Daemonum* ('The Devil's Signs') of 1563 suggested that mental illness, with its ravings and imaginings, was often and easily mistaken for witchcraft.

Were these disordered women – and they were mainly women – dangerous witches, needing to be burnt at the stake, or brain-sick lunatics? Regular physicians such as William Harvey (1578–1657), discoverer of the circulation of the blood and physician to King Charles I, had another role as detectors of witchcraft, advising courts on what was illness and what was not. His contemporaries included Richard Burton, author of the *Anatomy of Melancholy* (1621), an enormous compendium of exemplary cases – troubled minds, not necessarily 'depressed', in all their different lifestyles and presentations – and Richard Napier (see Table 1.2), an astrologer-cum-psychotherapist who gained a reputation for helping mentally ill people and whose notebooks still survive.

By about 1700 most of Europe was over its witch beliefs, and during the seventeenth and eighteenth centuries a rising demand for sympathetic treatment rather than incarceration demonstrated the swing from religious supernaturalism to rationalist and early scientific explanations. Now come the first medical writings on mental illnesses and the beginnings of the 'mad-doctoring' business. William Battie (1703–76), with his *Treatise on Madness* (1758) and a famous debate with John Monro (1715–91) as to the real basis of mental illness (deluded imagination or impaired judgement?), and George Cheyne (1671–1743), with his 1733 description of *The English Malady*, were the most famous of their time, doing very well for themselves in a seemingly lucrative trade. The better-off English middle classes had discovered neurosis.

Apart from a 1744 Vagrancy Act, however, the first law passed 'for the better maintenance and care of Lunatics' was in 1808, fear of mental illness creating shame and family cover-ups. When George III went mad in 1788, putting the constitution itself at peril, psychiatry was represented by a self-made Lincolnshire reverend/therapist, Francis Willis. The king was 'knocked flat as a flounder' via a behavioural regime that eventually brought him back to his senses, and he was even 'cured'. Treatment of disease rather than punishment for sins gradually became the standard approach. Contemporary descriptions of Parkinson's disease (in 1817), delirium tremens (in 1813) and the brains of post-mortem syphilitics (in 1822) demonstrated the physical basis of some (but not most) disordered behaviours.

THE RISE OF THE ASYLUM 1800–1900

The early private madhouses have few records, as privacy was required, but business flourished in the rising prosperity and urbanization of the eighteenth century. There had always been St Mary Bethlem (founded in 1247) – home of the Tom-of-Bedlams, wandering madmen, of myth and Shakespeare – and from the 1750s new charitable institutions (e.g. St Luke's in London, the Bethel in Norwich) were set up. A flourishing area for private madhouses was around Shoreditch, Bethnal Green and Hackney in east London, some of them family businesses ('Mr Balme's House') with lurid reputations. Concerns about maltreatment (patients left tied up in filthy, faeces-coated cots), and especially the death of Hannah Mills, in the York Hospital, led the local Quakers to set up, for their 'brethren', 'the Retreat', which became a European model. Not least this reputation was thanks to the masterly public relations of the 1813 *Description of the Retreat* by Samuel Tuke, outlining a form of treatment called 'moral therapy' based on kindness, reason, behavioural limits, exercise and activities, a family atmosphere and setting a good example.

In the year of the Battle of Waterloo (1815), there was also a widely reported Inquiry into Bethlem Hospital, detailing abuses and negligence, for example keeping one notorious patient in a metal contraption and chains for over 7 years (Figure 1.1). Especially scapegoated was the apothecary John Haslam, whose case study (1810) of Bethlem inmate James Tilly Matthews, *Illustrations of Madness* (see Table 1.3), is often quoted as the first description of a full-blown schizophrenia-type illness, with delusions, hallucinations and a bizarre range of other symptoms, including the patient's belief in an infernal air loom machine, run by a dangerous gang, that could do things such as 'thigh-talking', 'thought-making' and 'cutting soul from sense'. The ignorance of such disorders among the medical profession was a particular target of Haslam, since two prominent society physicians actually tried to have Tilly Matthews released, on the grounds that he was perfectly sane.

The resultant evangelical reform movement, led by the long-lived Lord Ashley, seventh Earl of Shaftesbury (1801–84), based on the principles of moral therapy, unchaining the insane (as Philippe Pinel in revolutionary Paris) and the

Figure 1.1 The iconic image of the Bethlem Inquiry: William Norris in the chained contraption in which he was kept in for 7 years

central role of the medical profession, created a specific therapeutic device for curing madness. This was the asylum, where 'lunatics' could be taken away from the hurly-burly of modern life (deemed to be driving them mad) and restored amidst fresh air, nice views, clean water and regular exercise. By 1845 every county in England had a statutory duty to build one (the Asylums Act). Such ring-fencing of money, for a particular specialty, has even been considered a kind of precursor to the start of the National Health Service (NHS), and over 120 asylums were eventually constructed in Britain (Figure 1.2), amidst a relentless rise in people deemed officially to be insane.

The psychiatric profession, throughout Europe, developed alongside the new asylums. Journals were founded, reports were published, posts were created and formal legal powers (certification) were introduced. Still readily stigmatized as 'mad-doctors' – hence the increasing fears about sane people being locked up for nefarious reasons (e.g. in novels such as *The Woman in White* by Wilkie Collins (1860) and *Hard Cash* by Charles Reade (1863)) – this new breed of 'alienists'-cum-'medical psychologists' wrote and held meetings and lobbied Parliament, but treatment approaches remained basic. Apart from the traditional purging, bleeding, confinement and exercise, they had to resort to devices such as whirling chairs, mustard baths (hot and cold), leeches and even early forms of electricity. Laying open the scalp, filling the wound with various herbs and ointments, and letting it suppurate for a week or two was practised by Dr Prichard, author of the much admired *A Treatise of Insanity*, published in 1835. For sedating medication they

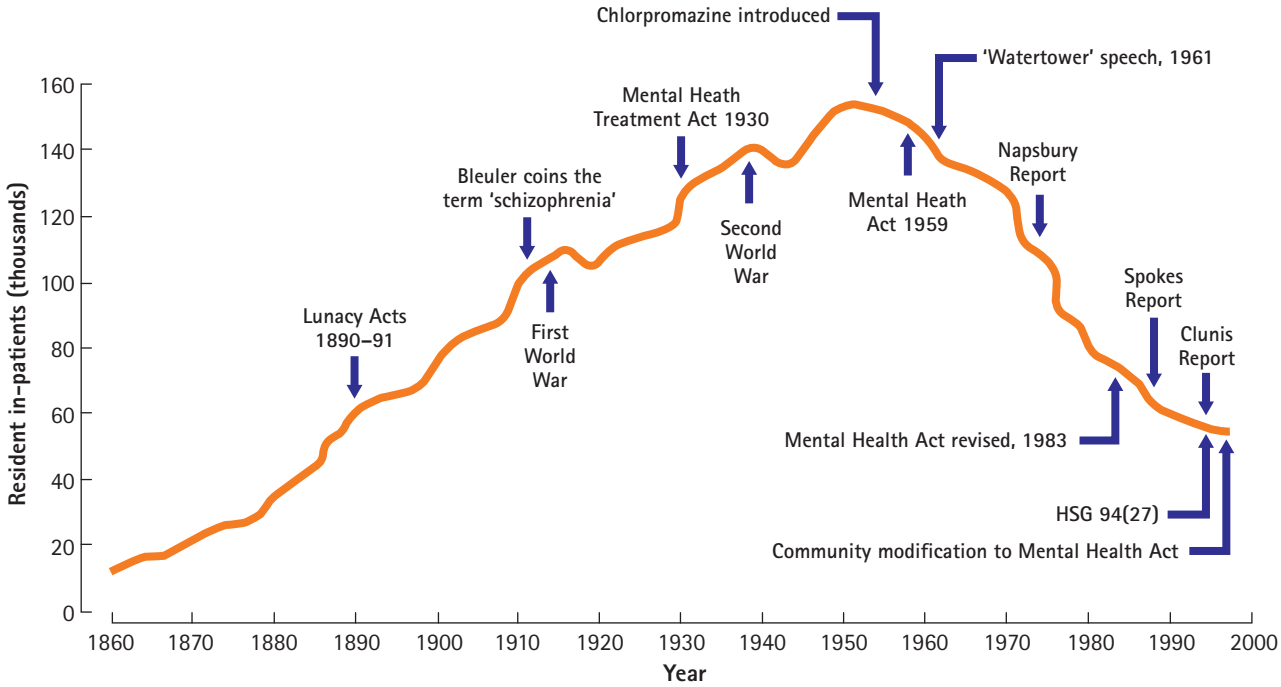

Figure 1.2 In-patients resident in mental illness hospitals, England and Wales, 1860–1995

had only opium/morphine, chloral hydrate and, of course, cannabis indica and brandy (Table 1.5).

But these first-generation psychiatrists and asylum superintendents did have one advantage. They could watch diseases evolve as the lunatic asylums grew larger and larger (see Figure 1.2), and they could start to classify the diseases more clearly. From the hordes of deranged, crippled, feeble-minded, epileptic and dementing inmates that filled up every new asylum, and each newly built asylum extension, doctors began to sort out different conditions, psychiatric, neurological, metabolic (e.g. hypothyroidism) and nutritional. Every leading alienist would publish his system, with new words, new diagnoses and new theories of causation. Abandoning the basics of 'mania' and 'melancholia', they created moral insanity, neurasthenia, catatonia, hebephrenia, lactational insanity and numerous others. By 1892 Daniel Hack Tuke, from the same family of Quaker Tukes that had founded the Retreat 100 years earlier, was able to publish his two-volume *Dictionary of Psychological Medicine*. This defined and discussed several thousand terms and phrases used by his zealous European-wide colleagues (including 200 legal cases).

But the asylums themselves, with their crowded 100-bed wards, their poorly paid and often untrained attendants, their handful of 'medical men' – perhaps four or five for two or three thousand patients – and their farms, drains and outbreaks of dysentery and typhoid, became less and less therapeutic and more and more just 'museums of madness'. Energetic superintendents burned themselves out trying to activate, morally treat, feed and semi-medicate their shuffling collections of 'chronic dements'. Paperwork and certificates dominated their days and nights, and ruthless magistrates imposed admissions and economies. Doctors could not refuse admissions, since all 'lunatics' were certified by the courts and simply sent in without warning. Formal inspections, unpleasant sounds and smells (incontinence was rife), and even random assaults dominated their day, although by the 1900s football matches and cricket leagues were well established. The good and bad of asylum life is summarized in Table 1.6.

Even worse news was that the numbers of insane people were rising faster than the population was growing, leading to fears of 'degeneration' – a theory first espoused by BA Morel (1809–73), a leading French asylum physician – and even a qualitative decline of the human race. Further classifying people as 'feeble-minded' led to more building (the Binet-Simon tests of intelligent quotient (IQ) appeared in 1905) of 'colonies' (i.e. very large asylums) of 'pauper imbeciles' designed to segregate the degenerate and to stop them overbreeding. In the early twentieth century some countries even sterilized large numbers of the 'mentally deficient', with certain American states such as California leading the way. The rise of eugenics (i.e. the theory of selective breeding of healthy genes) and the 1930s Nazis' attitudes to psychiatric patients, eliminating them as a precursor to the Holocaust, were the dark outcome of these developments.

The most positive effect of the asylums was the work of first-class researchers such as Emil Kraepelin (1856–1926). Using a computer-like system of library cards, he classified symptoms and followed their course, separating out dementia praecox (DP), a progressive, deteriorating psychotic disorder, from the better prognosis of manic-depressive illness, from which patients seemed to recover. This distinction, despite many attempts at redefinitions, remains at the heart

Table 1.5 Nineteenth-century medications

Traditional	Laxatives to 'purge' the bowel, e.g. croton oil
	Tonics, e.g. quinine, caffeine, strychnine
	Opium/morphine (hypodermic introduced 1855), cannabis, digitalis
1850s	Bromides
1870s	Hyoscyamus → hyoscine (from henbane plant) (also in combination with morphine/atropine)
1869	Chloral hydrate – widespread use in asylum, at home and as an addictive agent
1890s	Apomorphine – sedation and vomiting
1900s	Barbiturates

NB: By 1900, patent remedies were 72 per cent of pharmaceutical sales, often containing alcohol and stimulants (e.g. cocaine).

Table 1.6 Asylums: the good, bad (and ugly)

Positive	Negative
Sanctuary for 'lunatics'/ public safety	Stagnant, cut off from the 'normal' world
Food, shelter, clothes, 'attendants'	Internal abuses, neglect, scandals
Diagnostic research/ classification	Institutionalization*
Occupational activities (farm, laundry, woodwork)	Failure of moral therapy
Recognition of mental illness/ vulnerability	Distant, 'brooding', stigmatized
Financial/social commitment	Oversized, underfunded

*A twentieth-century term describing the effects on the individual of living in a controlled environment, with food, activities, contacts, clothes and time completely regimented by the needs of the institution.

of modern diagnosis, despite the term 'schizophrenia' replacing 'DP' as the standard terminology by the 1950s, probably because of its connotations of understandable treatability.

THE TWENTIETH CENTURY: I – THE BIOMEDICAL MODEL

Research advances in the late nineteenth century, especially in bacteriology, even started to pay dividends for psychiatry. Some 10 per cent of inmates of asylums had GPI, namely syphilis of the brain. When Noguchi and Moore discovered the syphilitic organism in 1913 (the brain abnormalities of syphilis had been known since the 1820s), it was the first clear connection of a mental illness to a disease process. Linked in with nineteenth-century knowledge of the effects of trauma to the brain, nutritional deficiencies, thyroid disease ('cretinism'), and the association between forms of epilepsy and types of hallucinations, a stronger line of research developed as to mental illness having a physical basis. Forget moral theories, 'religious excess' and the supernatural; here was cerebral disease generating crazy behaviours. Unfortunately this concentration on brain disorder did not generate much in the form of a biochemical marker or an effective treatment. Critics, often psychologists and psychotherapists, used the mocking term 'brain mythology'.

The pre-1950s asylums were full of chronic psychotics (schizophrenics and manic depressives), but the neuroses such as anxiety, depression, obsessionality and sexual perversions also had their proposed physical associations. A range of drugs, operations and shock therapies, of varying scientific reliability, ensued. Numerous records, photographic and written, graphically describe what happened. These unfortunately mainstream remedies may seem comic or gothic, for example excising colons or adenoids in search of 'focal infection', or giving injections of monkey gland cells to boost libido, but they reflect the desperation of patients, patients' families and physicians. The asylums had become enormous, crowded 'Private Worlds', as illustrated by Phyllis Bottome's 1932 novel of the same name and by the 1948 Warner Brothers movie *The Snake Pit*. In 1944, the textbook *Physical Methods of Treatment in Psychiatry* (Sargant and Slater) included insulin treatment, convulsion therapy (i.e. ECT), chemical sedation with drugs such as bromides and stimulation via Benzedrine (amphetamine), narcotherapy, malarial treatment (for GPI: give the patient malaria and a high fever and kill off the bugs before killing off the patient) and prefrontal leucotomy.

In the 1950s the discovery of chlorpromazine, other antipsychotics and a range of antidepressants led to the psychopharmacological revolution (Tables 1.7 and 1.8) and even new diagnoses to fit the new formulations available. The rise of the international drug companies in the 1970s and 1980s, pushing products such as Prozac, became a stock market phenomenon. Such approaches also enabled the gradual closure of the asylums, towards what has been called 'community care', and an increasing sense of neglect in that the sanctuary provided by the old asylums was no longer available. The continued failure, despite billions of dollars spent, to establish a clear biological marker for schizophrenia, depression or any other mental illnesses (although some brain imaging techniques are showing promise) has created a semi-mythical feeling about the need for 'counselling'. But the drugs do work, thus the now vociferous user movement and a rising anti-stigma campaign.

Table 1.7 Treatments for schizophrenia/psychosis

Pre-1950s	Sedative: chloral, hyoscine, opiates
	Physical: insulin coma, leucotomy, prolonged sleep
	Chemical/electrical shock
	Surgical excision (tonsils, colon, teeth)
	Psychological: psychoanalysis in various forms
1952	Deniken and Delay report on use of chlorpromazine in 38 patients treated at St Anne's Hospital, Paris
1952–1960	Widespread research/use worldwide; asylum numbers start to fall; term 'neuroleptic' coined
1958–1960	Haloperidol introduced
1960–1970s	Depot preparations of various dopamine-blocking antipsychotics
1969	Sulpiride
1974	Clozapine (reintroduced in the late 1980s)
1990s	'Atypical antipsychotics'

THE TWENTIETH CENTURY: II – PSYCHOLOGICAL APPROACHES

The asylums and classifications of the nineteenth century also led to a richer language of psychological description and a rising interest in psychological methods of treatment. Dominant personalities (among many) were Sigmund Freud (1856–1939) and Pierre Janet (1859–1947), the former's *Interpretation of Dreams* being published (but largely unsold) in 1900. For reasons that are hard to understand – although playing the sex card helped? – Freudian psychoanalysis swept aside many of its rivals. Like the Bolsheviks

Table 1.8 The road to modern psychopharmacology

1855	Von Bibra identified 17 different mind-altering plants (*Die Narkotischen Genusmittel unde der Mensch*)
1894	Adrenaline identified
1898	'Mezcal: a new artificial paradise' in *Contemporary Review* (a magazine of the time) by Havelock Ellis – 'an afternoon with peyote most educated gentleman should try once or twice'
1901	'Adrenalin' preparation marketed
1926	Acetylcholine identified in the synapse by Henry Dale
1929	Amphetamine (β-phenyl-isopropylamine) first used – 'feeling of wellbeing' noted 18 min after injection
1943	Lysergic acid diethylamide (LSD) discovered by Albert Hofmann – while cycling home, he found 'buildings yawned and rippled' and that his neighbour was wearing 'a lurid mask'
1950	Chlorpromazine (4560RP) and meprobamate synthesized
1955	Reserpine and meprobamate introduced for anxiety/depression
1956	Methylphenidate 'the happy medium in psychomotor stimulation' (marketed, and later targeted at childhood problems in 1961)
1957	Chlordiazepoxide, from a class of chemical dyes, 'causes a mouse to hang limply when held' (marketed in 1960 as Librium)
1957–1960	Imipramine, isoniazid and amitriptyline introduced as specific antidepressants
1959	Diazepam (marketed as Valium in 1963)
1974	Laboratory testing of fluoxetine (Lilly 110140)
1981	Zimelidine the first selective serotonin-reuptake inhibitor (SSRI) to be marketed (withdrawn in 1983 due to neurotoxic side effects)
1987	Fluoxetine (Prozac®) approved as an antidepressant
1994	Prozac had become the second best-selling drug in the world

in revolutionary Russia, the diehard commitment of Freud's followers created a new language of 'id', 'ego', 'transference' and 'Oedipus complex', loved by intellectuals and fitting the zeitgeist of post-First World War Europe. The shell shock of tens of thousands of previously 'normal' young men also helped to strengthen beliefs that mental illness was not due to heredity or degeneration, but rather was the understandable effects of a hellish environment, trench warfare. The treatment of the poet Siegfried Sassoon by the British psychoanalyst WHR Rivers (1864–1922), at Craiglockart Hospital, is especially well documented.

Although basing his theories upon some half a dozen famous cases, such as the Wolf Man (Sergei Pankaev, who died in 1979 having wandered around Europe, uncured, looking for help from many other therapists), Freud's approach to individual psychotherapy became the shibboleth of the cultured middle classes, especially in the USA. Despite schismatic alternatives developing under the leadership of Carl Jung (1875–1961) – the founder of analytical psychology and theories about archetypes – and Alfred Adler (1870–1937) – inventor of the inferiority complex, elaborated in *The Nervous Character* (1912) – the notion of the psychiatrist as a German Jewish professorial figure who put you on a couch while listening silently became the standard image. Even so, an undercurrent of – often misinformed – critical attitudes towards psychoanalysis persisted in mainstream medicine, not helped by some of the cruder approaches about childhood sexuality made by Freudian disciples such as Ernest Jones, who had to go off to Toronto to get away from the paedophilic taint. Scepticism also met the Freudians' beliefs that they could even 'cure' serious psychoses such as 'schizophrenia'. Given the looseness of that diagnostic term (i.e. the mistaken notion of a 'split' mind, like Dr Jekyll and Mr Hyde), this was not surprising. It took several decades of careful worldwide research to develop a reliable 'operational' system of diagnosing schizophrenia during the 1960s and 1970s, helped by the rise of effective antipsychotics, the need to counter the criticisms of anti-psychiatry and the new brain scanning techniques, such as magnetic resonance imaging (MRI).

In those same decades, however, radical psychology (e.g. RD Laing, Thomas Szasz), alongside alternative anti-establishment cultures, reached its pinnacle. This became the dominant mode of understanding the world, was critical of hidebound, asylum-bound, biology-bound 'straight' psychiatry, and scored notable victories. Diagnostic categories

were seen as loose or corrupt (e.g. the Russian KGB's term 'latent schizophrenia' for political dissidents), and mental illness was seen as caused by parental impositions (e.g. the 1971 film *Family Life*; Table 1.9), social injustice and restrictive mores. Therapeutic drug experience using lysergic acid diethylamide (LSD) was even mooted as the real breakthrough. Increasingly influential, however, were the behavioural psychologists, developing sophisticated theories of – and treatments for – anxieties and phobias, and moving on to cognitive behavioural therapy (CBT) and cognitive analytical techniques. Linking the psychoanalytical and behavioural approaches were group therapies, family therapies and patient-centred approaches such as Alcoholics Anonymous (AA) and an increasingly vociferous user movement.

Psychiatrists, from being asylum keepers in the nineteenth century to being proto-neurologists in the first half of the twentieth century, were now trying to be eclectic – that is, doctors-cum-therapists-cum-psychopharmacologists, with sophisticated interpersonal and therapeutic skills and an enhanced awareness of social and legal constructions. The organizations, publications and legislation psychiatry in Britain are summarized in Tables 1.10 and 1.11.

Table 1.10 Psychiatry in Britain

	Organizations
1841	Association of Medical Officers of Asylums and Hospitals for the Insane
1865	Medico-Psychological Association (MPA)
1926	Royal Charter → R(oyal) MPA
1971	Royal College of Psychiatrists
	Publications
1843	*Annales Medico-Psychologiques*
1844	*American Journal of Insanity*
1848–1854	*Journal of Psychological Medicine and Mental Pathology*
1853	*Asylum Journal* (ed. JC Bucknill)
1858–1962	*Journal of Mental Science*
1963	*British Journal of Psychiatry*
1970	*Psychological Medicine*
1977	*Bulletin of the Royal College of Psychiatrists* (→ *Psychiatric Bulletin* 1988)

WHAT NOW?

By the 1990s, the 'talking therapies', although intellectually established in terms of art criticism and literature, had run into a series of debates as to effectiveness, cost–benefit and the nature of 'distress'. Why talk à la Woody Allen to a shrink three times a week for 15 years, when Prozac could make you feel alright within a month? Newly recognized,

Table 1.9 Psychiatry's history in the cinema

Witchfinder General **(1968, dir. Michael Reeves)**	'In 1645 a villainous lawyer finds it profitable to travel the country instigating witch hunts'. Hammer horror, colourful and savage, absorbs you in its world of supernatural belief and religious terror. The 'witchcraft delusion' as reality
Freud **(1962, dir. John Huston)**	Earnest and hagiographic biopic set in 1885 Vienna, when young Dr F uses hypnotism and discovers a boy who hates his father because he loves his mother; dream sequences and detailed discussions follow. A perfect outline of mid-century certainties as to the 'science' of psychoanalysis
The Snake Pit **(1948, dir. Anatole Litvak)**	Set in a 1940s' New York asylum, this film depicts electroconvulsive therapy (ECT) without anaesthesia, crowds of caged patients, chain-smoking psychiatrists and a heroic Freudian doctor/analyst. Warnings were posted that children were not allowed in the cinema and that conditions in British mental hospitals were 'unlike' those depicted
One Flew Over the Cuckoo's Nest **(1975, dir. Milos Forman)**	Set in a Californian state mental hospital, with staff and patients playing some roles, this reflects 1950s' rather than 1970s' in-patient treatments, but it is the classic 1970s' anti-psychiatry, patient-as-freedom-fighter (Jack Nicholson) movie, with 'Big Nurse', ECT and lobotomy used to enslave the inmates. A powerful reinforcement of the continuing stigmatization of psychiatry, and psychiatrists, in the movies and the media
Family Life **(1971, dir. Ken Loach)**	A small, TV-style, British take on a 'schizophrenogenic' family driving a young woman into psychosis via double-binds, excessive criticism and suburban life. Vividly redolent of the late 1960s and the predominant psychosocial theories

Table 1.11 Legislation (England and Wales)

1808	Act for 'better maintenance and care of Lunatics'
1844–5	Lunatics and Asylum Acts – statutory duty to provide asylums; lunacy commissioners appointed
1890	Lunacy Act – legalism triumphant, complex certification process
1913	Mental Deficiency Act – 'colonies' for mental defectives; 'Board of Control' to supervise asylums
1930	Mental Treatment Act – out-patients and involuntary admissions allowed
1959	Mental Health Act – ending of magistrates' courts and Board of Control; review tribunals introduced
1983	Mental Health Act – introduced consent, approved social workers (ASWs) and a Mental Health Act Commission
1995	Patients in the Community Act – Supervised Discharge and Supervision Register (abandoned later)
2007	Revised Mental Health Act – approved clinicians and approved mental health professionals

though still controversial (vis-à-vis their aetiological role), conditions such as child sex abuse and post-traumatic stress disorder were rife, but talking about them didn't necessarily help. 'Counselling' research in primary care showed limited evidence of making you feel better or even happier; CBT has become almost a panacea.

By the end of the twentieth century, therefore, psychiatric practice had become an embattled profession. Still enmeshed in stigma, and beset by homicide inquiries, risk management and a rising demand for public safety, psychiatrists are seen by many as poodles of corporate power (the drug companies) failing to deliver enough talking therapy or real understanding of user problems. Their own history has even been subverted by historical sociologists, who view its so-called 'advances' as nothing more than a cover-up for Orwellian social control and the medicalization of unhappiness. Ever the war zone of conflicting attitudes to the normal and the mad, psychiatry therefore continues to find itself fighting for the neglected Cinderellas of society, nowadays the prison-housed thousands yearning for understanding (asylum and prison populations tend to relate inversely, the one rising as the other falls, and vice versa). Key themes of this historical process are outlined in Table 1.4.

This 'history of psychiatry' is merely one version of a complex series of events, discoveries, social movements and intellectual developments. It has barely touched upon the philosophical basis to modern psychopathology, the numerous Parliamentary committees generating legislative change, the popular books and articles informing public attitudes, or even the individual case histories that have influenced clinical practice. There is a rich range of sources still awaiting thoughtful research, and some degree of cynicism as to, for example, why doctors became the agreed experts on managing mental illness. But one theme is constant: psychiatrists are in demand because of their unique combination of clinical and personal skills. History enables us to know what those mean and how they can work.

Chapter 2

Introduction to evidence-based medicine

Stuart Carney

INTRODUCTION

It is hard to argue with the principle that decisions should be informed by the best available evidence. However, since the early 1990s, evidence-based medicine (EBM) has evoked fierce debate. This debate has focused essentially on two questions: what is the best available evidence for clinical decision-making, and what role should this evidence play?

In this chapter, I use the term 'evidence-based medicine' to refer to decision-making in the care of an individual patient. Other terms in common use are 'evidence-based healthcare' (EBHC) and 'evidence-based practice' (EBP). EBHC relates to decision-making about groups of patients or populations. The generic descriptor EBP is often used to embrace both EBM and EBHC.[1]

EBM grew out of the need to improve the effectiveness and efficiency of medical education. It offers clinicians, mindful of their limitations, a strategy for recognizing and managing clinical uncertainty and information overload. With over two million articles published in 20 000 biomedical journals each year, it is impossible for busy clinicians to keep up to date without competence in EBM.

THE BEST AVAILABLE EVIDENCE

EBM proposes a paradigm shift in clinical decision-making. Instead of relying on reasoning from basic sciences, EBM demands that clinical decisions are informed by the best available clinical research data. In other words, evidence showing that a drug *does* work in clinical practice is valued above inferences from basic sciences that a drug *should* work.

Implicit in this paradigm shift is the principle of a hierarchy of evidence. Just as clinical trial data trump reasoning from basic sciences when considering what drug to prescribe, a systematic review of two or more randomized controlled trials (RCTs) will typically provide more convincing evidence than an individual RCT, which again provides more convincing evidence than an individual cohort or case–control study. Recognizing the inherent strengths and weaknesses of different study designs for different types of question is essential for the efficient identification of the best available evidence. The development of electronic databases has made searching for and retrieving research data much easier.

As a reader of the biomedical literature, you will be aware that publication in a scientific journal is no guarantee of study quality. The EBM revolution serves to equip all healthcare practitioners with the opportunity to question received wisdom and constructively, yet critically, appraise research findings in order to determine whether they are valid, important and applicable to the specific clinical situation. Critical appraisal requires a working knowledge of the possible pitfalls in scientific research.

Having identified the best available evidence, what role should it play in decision-making?

THE ROLE OF EVIDENCE IN CLINICAL DECISION-MAKING

Dave Sackett, one of the earliest advocates, defined EBM as 'the conscientious, explicit, and judicious use of best evidence in making decisions about the care of individual patients'.[2] Critics of EBM have drawn attention to the limitations of using research evidence alone when making clinical decisions. Approaches such as values-based medicine and 'the expert patient' have challenged traditional models for the doctor–patient relationship.[3,4] As these models have evolved, so has the philosophy of EBM. Perhaps EBM can be best understood as a component of clinical expertise that is 'the ability to integrate research evidence and patients' circumstances and preferences to help patients arrive at optimal decisions'.[5] This model for patient decision-making is illustrated in Figure 2.1.

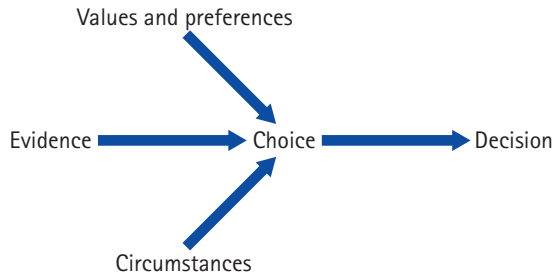

Figure 2.1 Role of evidence in patient decision-making

EVIDENCE-BASED PRACTICE AND CLINICAL EXPERTISE

Competence in EBM should help us empower patients to make optimal decisions. Helping patients to arrive at optimal decisions requires new and enhanced skills. These skills include the ability to use electronic databases, to critically appraise clinical research and data, and to communicate effectively with patients and other colleagues. Books and courses can help us develop our knowledge base, but the most effective way of developing competence in EBM is through reflective practice – that is, learning embedded in clinical practice.[6]

Drawing upon the public health disciplines of epidemiology and biostatistics, EBM seeks to simplify the process of accessing, appraising and applying research evidence. The EBM components of clinical expertise can be broken down into five steps:[7]

1. Translation of uncertainty to an answerable question.
2. Systematic retrieval of best available evidence.
3. Critical appraisal of evidence for validity, clinical relevance and applicability.
4. Application of results in practice (empowering patients to make clinical decisions).
5. Evaluation of performance.

The knowledge, skills and attitudes underpinning each of these five steps are summarized in Box 2.1.

It is clearly impractical to systematically retrieve and critically appraise the primary literature for every question that emerges in clinical practice. Meta-resources present pre-appraised summaries addressing some of the more common clinical questions, while evidence-based guidelines are becoming increasingly common. The art of EBM is judging, in partnership with patients, which questions are addressed adequately through guidelines, meta-resources and those that require appraisal of systematically retrieved primary evidence.[8]

Whatever the source of the evidence, the critical step in this process is the last one: the assessment of performance. This is easier to describe than to practise, but as we strive to improve patient safety and care we must not only evaluate our fidelity to the EBM process but also assess whether patients have been helped. The evaluation of performance remains the greatest challenge to truly practising EBM. Why? – because it requires a 'culture that is free of blame and encourages an open examination of error and failure'.[9]

Box 2.1 Knowledge, skills and attitudes underpinning the five steps of evidence-based medicine

To practise the five steps, clinicians should be able to:

- ***Translate their clinical uncertainty into answerable questions:***
  - Be able to assess patients and formulate a management plan.
  - Be aware of their own limitations and uncertainties.
  - Be motivated to seek guidance from published literature and colleagues.
  - Be able to translate these uncertainties into clinical questions.
- ***Systematically retrieve the best available evidence:***
  - Have knowledge and understanding of the resources available.
  - Have knowledge and understanding of how research is catalogued and strategies for efficient retrieval.
  - Have knowledge and understanding of the hierarchy of evidence.
  - Be able to effectively and efficiently access appropriate research evidence.
- ***Critically appraise the evidence:***
  - Have knowledge and understanding of basic epidemiology, basic biostatistics, basic qualitative methods, basic health economic methods and the principles of guideline/protocol development.
  - Be able to critically appraise primary research evidence and secondary sources, including guidelines.
  - Be able to determine whether the appraised evidence is applicable to this patient.
- ***Apply the results in practice:***
  - Be able to effectively communicate the strengths and weaknesses of the evidence, in a way that is respectful of the individual's circumstances and preferences, so that the patient is able to make an informed decision.
- ***Evaluate own performance:***
  - Have knowledge and understanding of the strategies to evaluate performance, including the importance of accurate, legible records, the role of electronic databases and the principles of audit.
  - Be able to evaluate their own performance and that of their team, and be actively engaged in developing strategies for quality improvement.
  - Be committed to monitoring performance (in a blame-free culture of learning).

REFERENCES

1. Gray J (2001) *Evidence-Based Healthcare*, 2nd edn. Edinburgh: Churchill Livingstone.
2. Sackett DL, Rosenberg WM, Gray JA, *et al.* (1996) Evidence based medicine: what it is and what it isn't. *British Medical Journal* **312**: 71–2.
3. Woodbridge K, Fulford K (2004) *Whose Values? A Workbook for Values-Based Practice in Mental Healthcare*. London: Sainsbury Centre for Mental Health.
4. Department of Health (2001) *The Expert Patient: A New Approach to Chronic Disease Management for the 21st Century*. London: Department of Health.
5. Guyatt G, Cook D, Haynes B (2004) Evidence based medicine has come a long way. *British Medical Journal* **329**: 990–1.
6. Coomarasamy A, Khan K (2004) What is the evidence that postgraduate teaching in evidence based medicine changes anything? A systematic review. *British Medical Journal* **329**: 1017–21.
7. Dawes M, Summerskill W, Glasziou P, *et al.* (2005) Sicily statement on evidence-based practice. *BMC Medical Education* **5**: 1.
8. Straus SE, Green ML, Bell DS, *et al.* (2004) Evaluating the teaching of evidence based medicine: conceptual framework. *British Medical Journal* **329**: 1029–32.
9. Halligan A, Donaldson L (2001) Implementing clinical governance: turning vision into reality. *British Medical Journal* **322**: 1413–17.

Chapter 3

History and philosophy of science

Bill (KWM) Fulford and Tim Thornton

INTRODUCTION

The psychiatrist and neuroscientist Nancy Andreasen has pointed out that a crucial characteristic of psychiatry as a research-led clinical discipline is the unique way in which it combines some of the most challenging scientific questions with equally challenging conceptual problems. Andreasen argues that the conceptual difficulties that we face have been driven increasingly to the top of psychiatry's practical agenda by the recent rapid pace of advances in the neurosciences.[1] These conceptual and empirical difficulties are evident in a wide range of contexts; examples include (i) disputes around the current revision processes for both the *International Classification of Diseases* (ICD) and the American Psychiatric Association's *Diagnostic and Statistical Manual*;[2] (ii) the contested shift in the UK and in many other countries in service delivery from psychiatry to a more multidisciplinary approach;[3] and (iii) growing pressure for a move away from professional-led mental health and social care services to a more equal partnership between professionals and service users and carers.[4]

In this chapter we set the conceptual difficulties surrounding current challenges of this kind in their historical context, with the aim of illustrating the extent to which the history and philosophy of science provide both context and direction for responding to the issues that we face today. It is natural to think of history and philosophy as perhaps subordinate to science as resources for improving service delivery in psychiatry. What we hope this chapter will illustrate is that the uniquely challenging nature of psychiatry, as a research-led clinical discipline, demands the resources of rigorous historical and philosophical work alongside, and as a full partner to, equally rigorous empirical methods.

Clearly, in a chapter of this length, we will not be able to cover the history and philosophy of science as it relates to psychiatry in any way comprehensively. Rather than giving an overview, therefore, we will adopt a case-study approach, exploring in some depth one particular strand in the history and philosophy of psychiatry that we believe is particularly relevant to current challenges. This strand concerns the role of both meanings and values as a complement to the brain sciences. The key protagonists that we will be considering are, respectively, one of the fathers of modern descriptive psychopathology, Karl Jaspers, and the Kantian philosopher of science Wilhelm Windelband, both writing around the start of the twentieth century.

The first section of this chapter focuses on Jaspers' work and in particular on his attempt to reconcile *meaningful* understanding with causal explanations in psychiatry. The second section examines Windelband's work, which focuses on the role of *values in understanding individuals*. Finally, we bring the story right up to date with recent developments in the counterparts of Jaspers' and Windelband's work, respectively phenomenology and what has become known as 'values-based practice'. This concluding section illustrates how these developments build on philosophical and empirical sources to provide practical approaches to decision-making in psychiatry that are directly complementary to the resources of the sciences.

JASPERS ON THE ROLE OF UNDERSTANDING MEANING IN PSYCHIATRY

Karl Jaspers was born in 1883. He studied first law and then medicine at university, graduating as a doctor in 1908 when he started to work as an assistant in the department of psychiatry in the University of Heidelberg. Kraepelin's influential textbook of psychiatry was in its eighth edition at this time. As a young man, Jaspers read widely, studying not only medicine but also psychology and philosophy, and it was the depth of his philosophical understanding that equipped him to make his unique contributions to psychiatry.

Jaspers and psychiatry's first biological phase

Jaspers was working at a time in the development of psychiatry very similar to our own, in that there were dramatic advances in the neurosciences of the day: the period has indeed become known as 'psychiatry's first biological phase'. Jaspers' professor in the Heidelberg department of psychiatry was Franz Nissl, a neurohistologist who discovered the dye that allowed the structure of nerve cells to be

clearly seen for the first time. Using this technique, he had shown that the neurohistological changes in general paralysis were different from the changes described by Alois Alzheimer in dementia. General paralysis was a degenerative dementia that had swept Europe after the wars of the late nineteenth century. It was shortly to prove to be a form of neurosyphilis. These were dramatic discoveries, therefore, and the young Jaspers was impressed with Nissl as a scientist. But when it came to *clinical* work, however, Jaspers was considerably less impressed.

Psychiatry at the turn of the century in Germany had moved out of the large institutions into university clinics. There was considerable resentment among the institutional psychiatrists that their discipline had been taken over by academic neuroscientists, whose knowledge of clinical psychiatry was scant, and whom they perceived as being under the spell of a crudely natural scientific model, epitomized by the German psychiatrist Wilhelm Griesinger's famous aphorism 'Mental illnesses are brain illnesses'.[5] Psychiatric researchers at the time, such as Griesinger, Alzheimer, Nissl, Carl Meynert and Theodor Wernicke, were actively searching for the neuropathological changes by which they believed the major psychoses could be characterized. And, given their success with general paralysis, hopes were high. Jaspers shared these hopes. But he believed that the underlying biological approach had been pushed too far. 'These anatomical constructions, however, became quite fantastic (e.g. Meynert, Wernicke) and have rightly been called "Brain Mythologies".'[5]

Jaspers and the *Methodenstreit*

Jaspers' reservations about what he perceived as the excessively natural-scientific approach to psychiatry were driven by his understanding of the philosophical debates about psychology in the late nineteenth century, the so-called *Methodenstreit*. This concerned whether the human sciences (*Geisteswissenschaften*) should try to emulate their far more successful cousins the natural sciences (*Naturwissenschaften*) or whether they should go their own methodological way. 'Positivists', including John Stuart Mill in England and both Auguste Comte and Emile Durkheim in France, argued that the human sciences were no different from the natural sciences. Others argued that the human or cultural sciences were different from the natural sciences, in terms of either the nature of their subject matter or their methodology, or both. The latter, in Germany, included Heinrich Rickert, Wilhelm Dilthey and Wilhelm Windelband, to whom we will return in the next section. Crucially for Jaspers, the German philosopher and sociologist Max Weber was among the latter camp.

Jaspers met Max Weber in 1909. He was invited to join Weber's elite intellectual circle, and he quickly became one of Weber's key intellectual antagonists. Jaspers thought of Weber as the 'Galileo of the human sciences'.[6] Although Weber believed that the human sciences involved a distinctive approach, he believed that sociology, his own discipline, was a hybrid subject, living partly within the natural and partly within the human sciences.

Jaspers and general psychopathology

Jaspers regarded psychopathology as Weber regarded sociology. It lay both within the natural sciences, pursuing abnormalities of brain functioning, and within the human sciences, pursuing the experiences, aims, intentions and subjective meanings of its patients. Of course, at a time when psychiatry was dominated by the 'brain mythologists', Jaspers' major aim was to bring psychiatry back within the ambit of the human sciences. He wanted to balance things up. In Weber's work, therefore, which in turn had drawn on the work of Dilthey, Windelband and Rickert, Jaspers saw things falling into place, and much of Weber's social theory – interpretation/understanding, *Evidenz*, ideal types and so forth – was to find its way into Jaspers' psychopathology. Some time later he wrote:

> My article of 1912 and this present book (1913) were greeted as something radically new, although all I had done was to link psychiatric reality with the traditional humanities. Looking back now, it seems astonishing that these had been so forgotten and grown so alien to psychiatry. In this way within the confines of psychopathology there grew a methodological comprehension of something which had always been present, but which was fading out of existence and which appeared in striking reverse, 'through the looking-glass' as it were, in Freud's psychoanalysis – a misunderstanding of itself. The way was clear for scientific consciousness to lay hold on human reality and on man's mental estate, his psychoses included, but there was an immediate need to differentiate the *various modes of understanding*, clarify them and embody them in all *the factual content* available to us.[5]

The period 1909–13 was a time of high output for Jaspers. He wrote papers on homesickness, hallucinations, pathological jealousy, phenomenology, and the need for both 'causal' (natural scientific) and 'understandable' (human scientific) connections in psychic life. We will discuss two papers published in this period below: 'The phenomenological approach in psychopathology'[7] and 'Causal and "meaningful" connections between life history and psychosis'.[8] But the culmination of this burst of output was that, in 1911, he was commissioned by the publisher Springer to write a textbook of psychopathology. It was thus that his *General Psychopathology* (*Allgemeine Psychopathologie*) appeared in its first edition in 1913.

Jaspers on subjective and objective

In 'The phenomenological approach in psychopathology',[7] Jaspers sets out his account of the role within psychopathology for a phenomenological approach. As we will explain shortly, phenomenology is, in Jaspers' view, a form of static understanding by contrast with what he calls

genetic understanding (see below). But both forms of understanding have in common that they are attempts to explore subjective, as opposed to objective, symptoms. What is the distinction between subjective and objective? Jaspers describes objective symptoms as follows:

> Objective symptoms include all concrete events that can be perceived by the sense, e.g. reflexes, registrable movements, an individual's physiognomy, his motor activity, verbal expression, written productions, actions and general conduct, etc.; all measurable performances... It is also usual to include under objective symptoms such features as delusional ideas, falsifications of memory, etc., in other words, the rational contents of what the patient tells us. These, it is true, are not perceived by the senses, but only understood; nevertheless, this 'understanding' is achieved through rational thought, without the help of any empathy into the patient's psyche.[7]

The distinction between what is available to rational thought and to empathy is important and one to which we will return. It helps to form a broader conception of what is objective than would generally be accepted today. But this in turn gives rise to a correspondingly narrower sense of 'subjective':

> Objective symptoms can all be directly and convincingly demonstrated to anyone capable of sense-perception and logical thought; but subjective symptoms, if they are to be understood, must be referred to some process which, in contrast to sense perception and logical thought, is usually described by the same term 'subjective'. Subjective symptoms cannot be perceived by the sense-organs, but have to be grasped by transferring oneself, so to say, into the other individual's psyche; that is, by empathy. They can only become an inner reality for the observer by his participating in the other person's experiences, not by any intellectual effort.[7]

Jaspers complains that a purely objective psychology leads 'quite systematically to the elimination of everything that can be called mental or psychic'.[7] In order to illustrate what he means, Jaspers refers to the assessment of a patient's fatigue through measurable performances where 'It is not the feeling of fatigue but "objective fatigue" which is being investigated'.[7] This suggests a contrast between objective and measurable aspects of physiology and the subjective aspect of what it is like to be or to feel fatigued. Assessing such subjective symptoms requires a kind of imaginative transference or empathy, which, thus, lies at the heart of psychopathology.

Jaspers on phenomenological understanding

Having drawn this distinction between subjective and objective, Jaspers goes on to characterize the aims of phenomenological psychopathology in the following terms:

> What then are the precise aims of this much-abused subjective psychology? ... It asks itself – speaking quite generally – what does mental experience depend on, what are its consequences, and what relationships can be discerned in it? The answers to these questions are its special aims... So before real inquiry can begin it is necessary to identify the specific psychic phenomena which are to be its subject, and form a clear picture of the resemblances and differences between them and other phenomena with which they must not be confused. This preliminary work of representing, defining, and classifying psychic phenomena, pursued as an independent activity, constitutes phenomenology.[7]

Thus, the key aim of phenomenology, or static understanding, is to identify the specific psychic phenomena that are to be its subject and to form a clear picture of the resemblances and differences between them and other phenomena. How does it set about this task?

> We should picture only what is really present in the patient's consciousness; anything that has not really presented itself to his consciousness is outside our consideration. We must set aside all outmoded theories, psychological constructs or materialist mythologies of cerebral processes; we must turn our attention only to that which we can understand as having real existence, and which we can differentiate and describe. This, as experience has shown, is in itself a very difficult task. This particular freedom from preconception which phenomenology demands is not something one possesses from the beginning, but something that is laboriously acquired after prolonged and critical work and much effort – often fruitless – in framing constructs and mythologies.[7]

A key aspect of the task of getting a very clear picture of psychological phenomena is thus to attempt to strip away theoretical descriptions or constructs and to get back to the things themselves, to use a slogan of the philosophical phenomenologist Edmund Husserl (1859–1938) (the precise nature of whose influence on Jaspers is contested). Of course, from a modern perspective, the aim of a theory-free approach has echoes of the aetiological-theory-free approach of the glossary to ICD-8 and its successors in both the ICD and DSM series of classifications. Quite how successful a genuinely theory-free approach can be in the face of arguments as to the essential theory-ladenness of data is a matter of debate (see Fulford *et al.*[9]).

So far, Jaspers' discussion has presented the difficulties of the phenomenological method rather than offering specific guidance as to how it is to be achieved. The most concrete account offered of that in this essay runs as follows:

> How then do we proceed when we isolate, characterise and give conceptual form to these psychic phenomena? We cannot portray them, or bring them before our eyes in any way that can be perceived by the senses. We can only guide ourselves by a multiple approach. We have to be led, starting *from the outside*, to a real appreciation of a particular psychic phenomenon by looking at its genesis, the conditions for its appearance, its configurations, its context and possible concrete contents; also by making use of intuitive comparison and symbolization, by directing our

observations in whatever ways may suggest themselves (as artists do so penetratingly) …

[T]he phenomenologist can indicate features and characteristics, and show how they can be distinguished and confusion avoided, all with a view to describing the qualitatively separate psychic data. But he must make sure that those to whom he addresses himself do not simply *think* along with him, but that they *see* along with him.[7]

Jaspers on genetic understanding

As remarked above, Jaspers distinguishes between two legitimate forms of understanding (of subjective phenomena): static understanding, which he also calls phenomenology, and genetic understanding. In 'The phenomenological approach in psychopathology', he characterizes the differences thus:

'Genetic understanding' [is] the understanding of the meaningful connections between one psychic experience and another, the 'emergence of the psychic from the psychic'. Now phenomenology itself has nothing to do with this 'genetic understanding' and must be treated as something entirely separate.[7]

What then *is* genetic understanding? A fuller discussion is given in another essay from this period, 'Causal and "meaningful" connections between life history and psychosis'.[8] Again, an important clue comes from the distinction mentioned above between rational connections and those that require empathy. Taking the case of understanding a speaker and understanding what the speaker has said, Jaspers comments:

The first important differentiation was made by Simnel, who showed the difference between the understanding of *what has been said* from understanding the *speaker*. When the contents of thoughts emerge one from another in accordance with the rules of logic, we understand the connections *rationally*. But if we understand the content of the thoughts as they have arisen out of the moods, wishes, and fears of the person who thought them, we understand the connections psychologically or empathetically. Only the latter can be called 'psychological understanding'. Rational understanding always only enables us to say that a certain rational complex, something which can be understood without any psychology whatever, was the content of a mind; empathic understanding, on the other hand, leads us into the psychic connections themselves. Whereas the rational understanding is only an *aid* to psychology, empathic understanding *is* psychology itself.[8]

This fits the earlier quotation in which the rational content of what a patient reports was characterized as an objective symptom. Equally, the rational interconnection between reports, their implications and so forth are also outside the domain of subjective psychology. But if empathy, in the service of genetic understanding, does not chart the logical or rational connection between one thought and another, then what is the basis of such psychological understanding? Jaspers' answer is not entirely clear. He says: 'We experience immediate evidence which we cannot reduce further nor base on any kind of other evidence… Meaningful connections are ideally typical connections. They are self-evident.'[8]

In other words, when there is a self-evident connection between one state and another, a connection that is typical although perhaps not always realized in practice, genetic understanding of it, through empathy, is possible. Realizing that one has won the lottery and thus solved all one's financial problems might lead, ideally and typically, to a state of happiness. Normally, to explain someone's sudden happiness, we might simply and sufficiently say that they have just heard that they have won the lottery. But, of course, in some unusual cases, that might not be a reason for happiness. The normal connection is basic – no more needs to be said – but it need not hold in all cases. It is ideally typical.

Thus, the relationship between static and genetic understanding is like this. The former articulates and vividly presents what it is like, for example, to have a sudden realization or what it is like to be in a state of happiness. It makes these kinds of state clear for further enquiry before the imposition of psychological theory. Genetic understanding adds to this the connection of how one state arises – ideally and typically – from the other. Such connections are shared empathically by psychological subjects, including psychiatrists and their patients.

Jaspers' characterization of static understanding has echoes in a more recent debate within the philosophy of mind between 'theory theory' and 'simulation theory' accounts of knowledge of other people's minds.[10] According to theory theory approaches, access to, and thus knowledge of, other people's minds is mediated by possession of a theory of mind. The theory is a body of deductively structured generalizations about the unseen causes of observable (speech and other) behaviour. This approach to the epistemological problem of how we can know about other minds thus dovetails with what are called 'functionalist' approaches to the ontological problem of what sort of things mental states are. Functionalism characterizes mental states in causal and functional terms, mediating between perceptual inputs and behavioural outputs. In other words, according to functionalism, mental states are akin to software states running on the brain as a computer. By characterizing types of mental state in second-order terms, functionalism aims to answer the problem of relating minds and bodies without simply reducing mental states to brain states. Theory theory approaches deploy broadly functionalist characterizations of what mental states are to explain, in addition, how we can have knowledge of them (in the case of other people).

Simulation theory, by contrast, explains knowledge of other minds not by possession of a theory of minds but merely by possession of a mind itself. The idea is that it is possible to have knowledge of another person's mental

states by imaginatively putting oneself into his or her predicament. One 'runs' one's deliberative processes 'offline', as it were. Simulation theory is thus a form of empathy. But, unlike Jaspers' account, there is no restriction to non-rational patterns of thoughts emerging from one another. Indeed, one of the key arguments for simulation theory and against theory theory is that rational connections lie at the heart of mental phenomena but are not reducible to or codifiable in any set of principles of good thinking that could thus form part of a theory of mind.[11,12]

The limits of understanding

Despite its centrality in Jaspers' conception of psychopathology and thus psychiatry, understanding has limits. One kind of limit is quite general and concerns its scope by contrast with natural scientific explanation. In a section of 'Causal and "meaningful" connections between life history and psychosis' called 'The limits of understanding and the universal application of explaining', he says:

> The suggestive assumption that the psychic is the area of meaningful understanding and the physical that of causal explanation is wrong. There is no real event, be it physical or of psychic nature, which is not accessible to causal explanation ...
>
> The effect a psychic state may have could in principle lend itself to a causal explanation, while the psychic state itself of course must be phenomenologically (statically) understood. It is not absurd to think that it might one day be possible to have some rules which could causally explain the sequence of meaningfully connected thought processes without paying heed to the meaningful connections between them ...
>
> It is therefore in principle not at all absurd to try to understand as well as to explain one and the same real psychic event. These two established connections, however, are of entirely different kinds of validity.[8]

The thought here seems to be this: Understanding and explanation do not have two distinct subject matters. Rather, the difference between them is one of method or of the kind of intelligibility that they deploy. As applied in psychiatry, they share the same subject matter: 'real events' or 'thought processes', in Jaspers' terms. These can in principle be charted in either way: either by looking to the law-like causal relations between them or by looking to the meaningful relations between them.

The idea that neural events might be susceptible to two distinct patterns of intelligibility was articulated by the US philosopher of mind Donald Davidson (1917–2003). On his model of the mind, Anomalous Monism, the very same events that comprise mental events and that – according to Davidson – stand in essentially rational relations also comprise physical events and can be subsumed under nomological or law-like causal explanations.[13] When described in mental property terms, however, there are no laws that fit them. Hence – qua instantiations of mental properties – they are *anomalous*. But there are laws that fit them that use their physical or neurological properties.

Given this broad view of the relation between understanding and explanation, one might expect the following asymmetry between them. Although every mental event is also a physical event, not every physical event is a mental event. There were no mental properties in the event of comet Shoemaker Levy 9 colliding with Jupiter in 1994, for example. That collision was not a mental event. Thus, one might expect Jaspers to say that, although every event that can be understood can also be explained, not every event that can be explained can also be understood. It is rather curious, therefore, that he actually says: 'there is no event which cannot be understood as well as explained'.[8]

Although he does not recognize that plausible general limitation on understanding – that non-mental events can only be explained, not understood – Jaspers does suggest a more specific local limit in the case of psychopathology. He believes that 'primary delusions' cannot be understood. To unpack that claim, we will now outline his taxonomy of delusions.

Jaspers suggests that delusions fall into two kinds: primary and secondary, or delusions proper and delusion-like ideas. Primary delusions fall into four further kinds. The first is mentioned in *General Psychopathology* almost in passing: delusional atmosphere. He says:

> with this *delusional atmosphere* we always find an 'objective something' there, even though quite vague, a something which lays the seed of objective validity and meaning... Patients feel as if they have lost grip on things, they feel gross uncertainty...[5]

To a person with schizophrenia, the world as a whole can seem subtly altered, uncanny, portentous or sinister. This general transformation prompts Jaspers to say elsewhere: 'We observe that a new world has come into being'.[5] There are then three further forms of primary delusion:

> *Delusional perceptions*. These may range from an experience of some vague meaning to clear, delusional observation and express delusions of reference...
>
> *Delusional ideas*. These give new colour and meaning to memory or may appear in the form of a sudden *notion* – 'I could be King Ludwig's son' – which is then confirmed by a vivid memory of how when attending a parade the Kaiser rode by on his horse and looked straight at the patient...
>
> *Delusional awareness*. This constitutes a frequent element particularly in florid and acute psychoses. Patients possess a knowledge of immense and universal happenings, sometimes without any trace of clear perceptual experience of them...[5]

In each of these cases, there is a deep change in the experience of the significance of features of the world. In the case of delusional perceptions, an experience is transformed. In the case of delusional ideas, the significance of a memory is transformed. In delusional awareness, a delu-

sional idea springs unbidden. But in all cases: 'All primary experience of delusion is an experience of meaning'.[5]

[T]he *experiences of primary delusion are analogous to this seeing of meaning*, but the awareness of meaning undergoes a radical transformation. There is an immediate, intrusive knowledge of the meaning and it is this which is itself the delusional experience.[5]

The key feature of primary delusions, however, is that they are un-understandable. While secondary delusions or delusion-like ideas are, in principle, understandable in the context of a person's life history, personality, mood state or presence of other psychopathology, primary delusions have a kind of basic status.

We can distinguish two large groups of delusion according to their *origin*: one group *emerges understandably* from preceding affects, from shattering, mortifying, guilt-provoking or other such experiences, from false perception or from the experience of derealisation in states of altered consciousness etc. The other group is for us *psychologically irreducible*; phenomenologically it is something final. We give the term '*delusion-like ideas*' to the first group; the latter we term '*delusions proper*'.[5]

As Andrew Sims says in a contemporary introduction to descriptive psychopathology:

[W]hen we consider the middle aged schizophrenic spinster who believes that men unlock the door of her flat, anaesthetize her and interfere with her sexually, we find an experience that is ultimately not understandable. We can understand, on obtaining more details of the history, how her disturbance centres on sexual experience; why she should be distrustful of men; her doubts about her femininity; and her feelings of social isolation. However, the *delusion*, her absolute conviction that these things are happening to her, that they are true, is not understandable. The best we can do is to try and understand externally, without really being able to feel ourselves into her position (*genetic empathy*), what she is thinking and how she experiences it [Sims 1988: 85].[14]

Thus, although Jaspers places empathic understanding (both static and genetic) at the heart of psychopathology and thus psychiatry, he also argues that some of the key phenomena that characterize psychopathology cannot be understood. They are un-understandable. If Jaspers is correct, then psychiatry has a fundamental limitation. We return to this point later, when we bring the story up to date with developments in modern phenomenology.

WINDELBAND ON THE ROLE OF UNDERSTANDING VALUES IN PSYCHIATRY

In this section, we consider a second strand of conceptual work from psychiatry's first biological phase, namely the role of values, which as a complement to the neurosciences has particular relevance to psychiatry today. As noted earlier, although Jaspers had important things to say about the role of meanings, he was less interested in the role of values. In the following passage, for example, he distances his conception of understanding meaning from forming or understanding value judgements:

It is a fact that when dealing with meaningful connexions as such we inevitably tend to value positively or negatively, while everything meaningless we merely value, if we do so at all, only in relation to something else. Thus the emergence of moral demands from resentment we may value as something despicable, whereas we value memory merely as a tool. In the *science* of psychology, however, we must strictly refrain from any such value judgement. Our task is merely to grasp the meaningful connexions as such and to recognize them.[8]

Thus, Jaspers separates understanding meanings from the kind of evaluation that properly involves the assessment of values. But another European philosopher of science, Wilhelm Windelband, did wish to stress the importance of values for a properly rounded scientific account. He did this via an account of idiographic understanding, which, as we will outline later, is itself at the forefront of contemporary thinking about psychiatry.[15]

Windelband and idiographic understanding

Wilhelm Windelband was a Kantian philosopher of science. He first introduced the distinction between 'idiographic' and 'nomothetic' in his rectorial address of 1894. Key components of the distinction between them are that it is a distinction of method not of subject matter, that it concerns treating events as unrepeated, and that it is a reaction against an overreliance on an essentially general conception of knowledge.

Windelband contrasts his own methodological distinction with one of substance, between natural sciences (*Naturwissenschaften*) and sciences of the mind (*Geisteswissenschaften*): 'I regard the dichotomy as unfortunate. Nature and mind is a substantive dichotomy ... not equivalent to a dichotomy based on modes of cognition.[16]

Such a distinction of substance is a hostage to the fortune of a metaphysical distinction of kind between mind and the rest of nature. In psychiatry, the interplay of both broadly psychological methods and neurology makes drawing such a distinction premature and unhelpful.

Windelband proposes, instead, a distinction that places psychology (as he understands it) and other natural sciences on one side and other disciplines, which in Germany at the time were called 'sciences of the mind' but which have a distinct method, on the other. This gives rise to a characterization of what he goes on to label 'idiographic', as follows:

[T]he majority of the disciplines that are usually called sciences of the mind have a distinctively different purpose: they provide a complete and exhaustive description of a single, more or less

extensive process which is located within a unique, temporally defined domain of reality.[16]

Idiographic understanding and the uniqueness of individuals

As first introduced, idiographic understanding concerns individual or unique cases. But, given that the distinction is supposed to be at the level of method and not substance, this is fixed not by the subject matter so much as how that subject matter is approached. This is made clearer in the following passage, in which the term 'idiographic' is first introduced:

> In their quest for knowledge of reality, the empirical sciences either seek the general in the form of the law of nature or the particular in the form of the historically defined structure. On the one hand, they are concerned with the form which invariably remains constant. On the other hand, they are concerned with the unique, immanently defined content of the real event. The former disciplines are nomological sciences. The latter disciplines are sciences of process or sciences of the event. The nomological sciences are concerned with what is invariably the case. The sciences of process are concerned with what was once the case. If I may be permitted to introduce some new technical terms, scientific thought is *nomothetic* in the former case and *idiographic* in the latter case.[16]

This suggests the following rough practical distinction: nomothetic approaches are those that chart law-like, or nomological, generalities. Their aim is to describe generalities. Idiographic understanding concerns individual cases described in non-general ways. Both are, however, forms of empirical enquiry.

Such a distinction fits modern psychological usage influenced by Gordon Allport (1897–1967), in which 'idiographic' is used to describe case-study-based qualitative research by contrast with quantitative cohort-based research (although whether Allport's use of nomothetic accords with Windelband's is a matter of dispute[17]). But although contemporary use in psychology suggests that idiographic and nomothetic forms of understanding are practically distinguishable but not fundamentally distinct, Windelband suggests that his distinction goes deeper. He relates it to a fundamental metaphysical divide:

> [T]his distinction connects with the most important and crucial relationship in the human understanding, the relationship which Socrates recognized as the fundamental nexus of all scientific thought: the relationship of the general to the particular.[16]

What Windelband means is this. Scientific understanding has, since the Greeks, concentrated on generalities. It has concentrated on laws of nature that govern many objects and events. But in so doing it has neglected the importance of understanding individual objects, events or people not as instances of general kinds but in their individuality. He says:

> The commitment to the generic is a bias of Greek thought, perpetuated from the Eleatics to Plato, who found not only real being but also real knowledge only in the general. From Plato this view passed to our day. Schopenhauer makes himself a spokesman for this prejudice when he denies history the value of a genuine science because its exclusive concern is always with grasping the specific, never with comprehending the general… But the more we strive for knowledge of the concept and the law, the more we are obliged to pass over, forget, and abandon the singular fact as such…[16]

Unique individuals and values

This raises a question – which arguably Windelband never answered satisfactorily – as to the nature of this individualistic understanding. What is it to understand an individual in essentially non-general terms? But he did give an important clue as to why he thought there was need for idiographic understanding. What we value when we value people or events is tied to their individuality, he argues.

> In opposition to this [general, nomothetic] standpoint, it is necessary to insist upon the following: every interest and judgment, every ascription of human value is based upon the singular and the unique… Our sense of values and all of our axiological sentiments are grounded in the uniqueness and incomparability of their object.[16]

Examining value judgements helps to reveal the fundamental importance of particular cases as opposed to general kinds in judgements. It suggests that there is an important role for clinical judgement aimed at reflecting the nature of individuals and their experiences. Windelband himself seems to have taken this to imply the need for a particular kind of individualistic judgement in which there is no implicit comparison – as there is with any general concept – with other cases. Such a judgement would be essentially particular or individualized.

Windelband's fellow neo-Kantian Heinrich Rickert also argues that there can be essentially particular or individualized judgements and that these are exemplified by value judgements. Unlike nomothetic accounts of, for example, the forces acting on bodies, which are described and explained in general terms, judgements about the value of things are individualized judgements.

> We are concerned here with the connection of objects with values; for a generalizing approach the objects are free of value-connection, they are exemplars, replaceable… This is what happens when we free the object of all connection with our interests – it becomes a mere exemplar of a general concept. An individualising approach is necessarily connected with the value-bound grasp of the object [*mit der wertverbindenen Auffassung der Objekte*]…[18]

But although Windelband's discussion supports the suspicion of subsuming individuals under categories and the role, instead, of a kind of individualized judgement, he does

not offer a very clear account of what form such an idiographic judgement might have. How precisely is a judgement supposed to reject an historical overemphasis on the general? In fact, there is reason to be suspicious of the suggestion that reflecting individuals needs a special kind of individualistic judgement.[19,20]

Thus, the lasting importance of Windelband's discussion is this: he reminds us of the importance of the individual as well as the general. He suggests that judgement aimed at understanding individuals in their own terms rather than instances of generalities can be an important aspect of a scientific understanding, albeit one that contrasts with essentially general statistical or law-like explanation. And he suggests that understanding of individuals can be importantly connected to value judgements. In the final section of the chapter, we will outline a more recent approach to reflecting the values of individuals that avoids Windelband's reliance on the contested notion of 'individualising judgement'.

MEANING AND VALUES IN PSYCHIATRY TODAY

In this section we outline current developments in psychiatry as they reflect respectively meanings (phenomenology and related disciplines) and values (values-based practice), both of which are philosophically derived resources for clinical practice that are complementary to the resources of the sciences. Again, we are unable to cover either of these in detail, but we include a list of suggested further reading at the end of the chapter.

Phenomenology and related disciplines

Although not prominent in much of Anglo-American psychiatry, there was a strong continuing tradition of work in phenomenology and related disciplines in continental Europe through much of the twentieth century. This tradition was indeed one of a number of important sources of the remarkable resurgence of cross-disciplinary work between philosophy and psychiatry that began in the 1990s in parallel with the dramatic advances in the neurosciences of that period.[21]

As a rich theoretical discipline, phenomenology has continued to develop strongly along with other disciplines broadly within the 'philosophy of mind', very much in partnership with research in the neurosciences; see, for example, a number of papers in the double special issue of *Philosophy, Psychiatry, and Psychology*, edited by Christoph Hoerl.[22] As in Jaspers' day, it has been crucial in this respect that careful analysis of the subjective content of experience is available as a complement to the findings of empirical disciplines, such as functional neuroimaging and behavioural genetics. Equally important, though, has been the clinical impact of phenomenology, and related disciplines such as hermeneutics and existentialism, through improved understanding of the subjective experience of mental disorder. We return in a moment to a particular application of phenomenology to the problem of understanding as raised by Jaspers' work. Examples of other clinical applications of this area of philosophy include the work of the Dutch philosopher Guy Widdershoven on improved decision-making in old-age psychiatry;[23] of the American philosopher and psychologist Steven Sabat on interpretation of language difficulties in Alzheimer's disease;[24] and of the Oxford philosopher of mind Katherine Morris on body dysmorphic disorders.[25]

Returning, then, to the problem of understanding as raised by Jaspers, recent work by psychiatrists and philosophers alike has challenged the assumption that this rules out meaning-laden understanding as an aim, at least. The Italian psychiatrist and phenomenologist Giovanni Stanghellini, for example, treads a middle ground between explaining and understanding schizophrenia. In his book of essays *Disembodied Spirits and Deanimated Bodies*, he has argued that some understanding of the experiences of sufferers of schizophrenia is possible on the hypothesis that they experience a threefold breakdown of common sense.[26] This involves a breakdown of three distinct areas: the ability to synthesize different senses into a coherent perspective on the world (coenaesthia); the ability to share a common world view with other members of a community (sensus communis); and a basic pre-intellectual grasp of, or attunement to, social relations (attunement). Stanghellini says: 'The philosophical kernel of my proposal is to show how all these dimensions of the phenomenon of common sense (coenesthia, sensus communis, and attunement) *are related to each other*'.[26]

But Stanghellini does not attempt to use these ideas to step wholeheartedly inside the world view of subjects with schizophrenia. Rather, breakdowns of these are postulated as clues to interpret the strange things that people with schizophrenia report. But a basic phenomenon remains: the inaccessibility of experiences and thoughts:

> Listening to a person affected by schizophrenia is a puzzling experience for more than one reason. If I let his words actualize in me the experiences he reports, instead of merely taking them as symptoms of an illness, the rock of certainties on which my life is based may be shaken in its most fundamental features. The sense of being *me* the one who is now seeing this sheet, reading these lines and turning this page; the experience of perceptual unity between my seeing this book, touching its cover and smelling the scent of freshly printed pages; the feeling that it is me the one who agrees or disagrees with what I am reading; the sense of belonging to a community of people, of being attuned to the others and involved in my actions and future; the taken-for-granted of all these doubtless features of everyday life, may be put at jeopardy.
>
> Although my efforts to understand, by suspending all clinical judgement, allow me to see these person's self-reports as a pos-

sible configuration of human consciousness, I must admit that there is something incomprehensible and almost inhuman in these experiences, something that makes me feel radically different from the person I am listening to.[26]

This suggests that understanding is an ongoing and ultimately unfinishable task. The clinician has to make a series of interpretative judgements taking broad account of the life of the sufferer from, for example, schizophrenia. Such judgements can help towards at least a partial understanding of the person as a whole while at the same time taking account of the vividly alien quality of their psychopathological experiences.

On this approach, interpretative judgements presuppose that, as it were, the basic unit of meaning is the life of the whole person. Attempts at individual interpretation of specific delusions can be guided by a more general framework that takes schizophrenia, for example, to involve a breakdown of common sense. But such an approach goes only so far and, once they have given out, the interpretation of individual experiences has to be replaced by a sometimes partial and shifting understanding of the person as a whole based on contextual judgements. As in other areas of psychiatry, there is no quick route to bypass the need for good and sensitive judgement and hence to the irreducible role of the individual. It is the link between the general and the individual as mediated by values that is illustrated by recent developments in values-based practice outlined below.

Values and values-based practice

As described earlier, although Jaspers rather dismissed values, it was the particular contribution of Windelband during psychiatry's first biological phase to show their importance in relation to the uniqueness of individual people. As with phenomenology, then, so values have sprung back into prominence in recent years as part of the new philosophy of psychiatry, alongside and as a complement to developments in the neurosciences. The most familiar aspect of the new prominence of values is of course ethics; but other examples of emerging disciplines include health economics (e.g. Brown *et al.*[27]) and decision analysis (Hunink *et al.*[28]).

Values-based practice is distinctive theoretically in that it is derived from both philosophical and empirical sources. At a practical level, it is the closest to Windelband's work in providing a complement to the generalized sciences:

- in the emphasis it places on the importance of the *diversity of individual values*, including the values of clinicians, researchers and managers as well as those of patients and carers; and
- in relying on a number of elements of good *process*, in particular specific and learnable clinical skills, to support balanced decision-making where values conflict.

It is because it is process- rather than outcome-based that values-based practice is most directly complementary to the sciences as a resource for clinical decision-making. Values-based practice, as we describe further below, is indeed in this respect directly complementary to evidence-based practice.[29] In this section, we describe briefly (i) the theory and empirical base of values-based practice, including its philosophical roots, and (ii) illustrative examples of recent policy, training and service development initiatives in values-based practice in the UK and internationally.

The theory and empirical base of values-based practice

The theory underpinning values-based practice is based on work in linguistic analytical philosophy of the 'Oxford school' in the middle decades of the twentieth century, on the meanings of key value terms, such as 'good', 'ought' and 'right'. Exemplar work from this period includes RM Hare's *The Language of Morals*,[30] *Freedom and Reason*[31] and *Descriptivism*,[32] in which he explored the logic of value terms. Although not drawing directly on the work of Windelband and others in the *Methodenstreit*, Hare can be understood as being concerned with broadly the same issues, namely how factual terms are related to value terms. Hare's line on this was that there is always a logical distinction (i.e. a distinction of meaning) to be made between these two kinds of term: this is essentially because to evaluate something as good or bad always means something more than merely describing it. Thus, in one of Hare's examples, an eating apple that is (i.e. can be described as) red and crisp happens to be, for most people, a good eating apple; but to actually call such an apple a 'good eating apple' means more than merely describing it as red and crisp – it also commends it.

Hare's work did not go uncontested, of course: alternative views were put forward, for example by another Oxford philosopher, GJ Warnock, in his *The Object of Morality*.[33] The debate between Hare and Warnock was itself set in a tradition of analytical philosophy running through much of the twentieth century,[34] and also back to the work of the British empiricist philosopher David Hume.[35] The debate indeed continues to this day; see, for example, the 2002 collection by the American philosopher Hilary Putnam.[36]

Nonetheless, work in this tradition, and in particular Hare's disentangling of descriptive and evaluative meaning, provides a powerful set of insights that, although not developed originally with medicine in mind, can help us to understand the relationship between facts and values in healthcare. Fulford applied these insights to the meanings of medical terms such as 'illness', 'disease', 'disability', 'function' and 'dysfunction' in his *Moral Theory and Medical Practice* (Fulford, 1989).[37]As a contribution to the continuing debate about the meanings of these medical terms as they are used particularly in psychiatry, it is Fulford's work, in *Moral Theory and Medical Practice* together with a number of subsequent articles,[38–40] that provides the key theoretical underpinnings for values-based

practice. We do not have space here to go into the theory of values-based practice in detail, but among the practical implications of the theory are clear ways of articulating the relationships between values-based practice and both ethics- and evidence-based practice.[29]

Values-based practice also builds on empirical work on values guided by the philosophical theory just outlined. Thus, a key prediction of the theory of values-based practice is that the implicit values driving medical decision-making are often far more diverse than is generally recognized. This prediction has been tested by the British social scientist Anthony Colombo in a major study of the models of disorder (including values and beliefs) guiding decisions in the management of people with long-term schizophrenia in the community.[41] The study combined empirical methods from the social sciences with the analytical philosophical theory just outlined and was innovative in a number of respects.[42] The results of the study were widely disseminated in both research and non-governmental organizations (NGO) journals (e.g. in the house journal of MIND), and the methods developed in this study became the basis for one of the main areas of skills training for values-based practice (see below).

Policy, training and service development initiatives in values-based practice

As noted above, values-based practice starts from the diversity of individual values and relies on good process to support balanced decision-making in practice. In its reliance on good process, rather than preconceived correct outcomes, values-based practice is very much a partner to evidence-based practice: evidence-based practice provides a process for effective healthcare decision-making where the relevant evidence is complex and possibly conflicting; values-based practice provides a different but complementary process for effective healthcare decision-making where the relevant values are complex and possibly conflicting.

Again, we do not have space here to describe the process of values-based practice in detail. Ten key principles of the process of values-based practice have been set out by Fulford, together with a detailed case history of a patient with manic–depressive disorder, showing how each of these principles interweaves in practice with evidence-based approaches.[29] The ten principles have been applied in mental health and social care as the basis of a series of policy, training and service development initiatives. This work has been carried out in partnership with both service users and service providers and with institutional support from NGOs, including the Sainsbury Centre for Mental Health (SCMH), the Mental Health Foundation (MHF) and Turning Point in London, and the World Psychiatric Association, and from government departments, in particular the UK Department of Health. Internationally, corresponding developments have been included in the World Psychiatric Association's Institutional Program on Psychiatry for the Person.[43,44] Examples of these developments are described below.

Policy

The National Institute for Mental Health in England (NIMHE), as the body responsible for mental health policy implementation in England and Wales, published a framework of values that was based explicitly on the approach of values-based practice (Box 3.1).[45] This in turn guided a range of subsequent policy developments, including a generic skills programme,[46] as the basis of moves towards more multidisciplinary and person-centred approaches to service delivery.[3] The approach has also been applied in a number of specific areas of policy development, including delivering race equality[47] and the introduction of community development workers.[48]

Training

The first training manual for values-based practice was developed in a partnership between the Sainsbury Centre for Mental Health and Warwick Medical School, with the support of NIMHE. Published as *Whose Values?*,[49] the training manual was launched at a conference in London by the minister Rosie Winterton and, together with the NIMHE values framework, supported the policy initiatives noted above. A further application of values-based practice has been in the training materials produced to support the amended Mental Health Act in England and Wales. These training materials combine evidence-based resources with an innovative values-based approach to using the guiding principles defined by the code of practice (Table 3.1) as a framework of values guiding the application of the general provisions of the Act to individual cases (Figure 3.1).[50] Outside the UK, training programmes in values-based

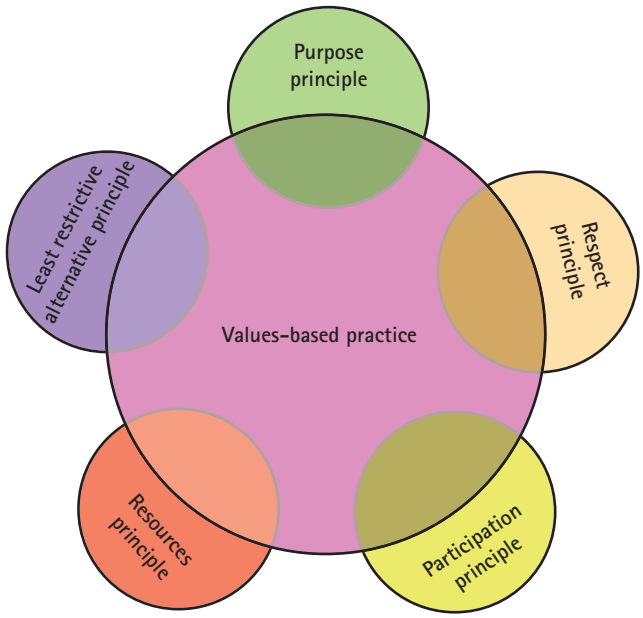

Figure 3.1 The guiding principles as a framework of values. Reproduced with permission from Ref. 50

Box 3.1 National framework of values for mental health

The work of the National Institute for Mental Health in England (NIMHE) on values in mental health care is guided by three principles of values-based practice:

- *Recognition:* NIMHE recognizes the role of values alongside evidence in all areas of mental health policy and practice.
- *Raising awareness:* NIMHE is committed to raising awareness of the values involved in different contexts, the role(s) they play and their impact on practice in mental health.
- *Respect:* NIMHE respects diversity of values and will support ways of working with such diversity that makes the principle of service-user centrality a unifying focus for practice. This means that the values of each individual service user/client and their communities must be the starting point and key determinant for all actions by professionals.

Respect for diversity of values encompasses a number of specific policies and principles concerned with equality of citizenship. In particular, it is anti-discriminatory because discrimination in all its forms is intolerant of diversity. Thus, respect for diversity of values has the consequence that it is unacceptable (and unlawful in some instances) to discriminate on grounds such as gender, sexual orientation, class, age, abilities, religion, race, culture or language.

Respect for diversity within mental health is also:

- *user-centred:* it puts respect for the values of individual users at the centre of policy and practice;
- *recovery-oriented:* it recognizes that building on the personal strengths and resiliencies of individual users, and on their cultural and racial characteristics, there are many diverse routes to recovery;
- *multidisciplinary:* it requires that respect be reciprocal, at a personal level (between service users, their family members, friends, communities and providers), between different provider disciplines (such as nursing, psychology, psychiatry, medicine, social work), and between different organizations (including health, social care, local authority housing, voluntary organizations, community groups, faith communities and other social support services);
- *dynamic:* it is open and responsive to change;
- *reflective:* it combines self-monitoring and self-management with positive self-regard;
- *balanced:* it emphasizes positive as well as negative values;
- *relational:* it puts positive working relationships supported by good communication skills at the heart of practice.

NIMHE will encourage educational and research initiatives aimed at developing the capabilities (the awareness, attitudes, knowledge and skills) needed to deliver mental health services that will give effect to the principles of values-based practice.

practice have been developed in a number of European countries and in South Africa.[51]

Service developments

Many of the above policy and training initiatives have been associated with corresponding developments in service delivery. An example of a direct application of values-based practice to service development is a programme on diagnostic assessment in mental health and social care. This programme followed a series of international research seminars, initially in Dallas, USA,[52] and subsequently in London, supported by the Department of Health, exploring the role of values in psychiatric diagnosis. The last of these seminars, which was hosted jointly with the Mental Health and Substance Abuse section of the World Health Organization (WHO), led to the launch of a wide-ranging consultation on approaches to assessment that were considered best practice by all stakeholders – that is, by service providers from different disciplinary backgrounds, and also by service users and carers.

The programme on diagnostic assessment was distinctive in building on a specific prediction of the theory of values-based practice that, in psychiatry, values are as important in diagnosis – that is, in how we come to understand a problem – as they are in what we do about it – that is, in such areas as treatment and resource allocation. The programme has strongly practical outcomes, however, much of the publication itself comprising examples of innovative practice from the field. As with other work in values-based practice, these examples illustrate in a very practical way how the resources of generalized evidence-based science can be combined with an approach to assessment that is fully responsive to the needs, wishes, strengths and other values of the individual concerned (see Box 3.2).[53]

CONCLUSIONS

The period around the start of the twentieth century was one of great progress for biologically based neuroscience. It was a time of great promise for the science of psychiatry.

Box 3.2 Three keys to a shared approach to assessment in mental health and social care

- *Active participation* of the service user concerned in a shared understanding with service providers and, where appropriate, with their carers
- Input from *different provider perspectives* within a multidisciplinary approach
- A person-centred focus that builds on the *strengths, resiliencies and aspirations* of the individual service user and identifies his or her needs and challenges.

Table 3.1 Guiding principles in the code of practice for the new Mental Health Act (England and Wales)

Purpose	Decisions under the Act must be taken with a view to minimizing the undesirable effects of mental disorder, by maximizing the safety and wellbeing (mental and physical) of patients, promoting their recovery and protecting other people from harm
Least restrictive alternative	People taking action without a patient's consent must attempt to keep to a minimum the restrictions they impose on the patient's liberty, having regard to the purpose for which the restrictions are imposed
Respect	People taking decisions under the Act must recognize and respect the diverse needs, values and circumstances of each patient, including their race, religion, culture, gender, age, sexual orientation and any disability. They must consider the patient's views, wishes and feelings (whether expressed at the time or in advance), so far as they are reasonably ascertainable, and follow those wishes wherever practicable and consistent with the purpose of the decision. There must be no unlawful discrimination
Participation	Patients must be given the opportunity to be involved, as far as is practicable in the circumstances, in planning, developing and reviewing their own treatment and care in order to help ensure that it is delivered in a way that is as appropriate and effective for them as possible. The involvement of carers, family members and other people who have an interest in the patient's welfare should be encouraged (unless there are particular reasons to the contrary) and their views taken seriously
Resources (effectiveness, efficiency and equity)	People taking decisions under the Act must seek to use the resources available to them and to patients in the most effective, efficient and equitable way, to meet the needs of patients and achieve the purpose for which the decision was taken

Nevertheless, two figures of great importance for psychiatry realized that the science of natural laws and general mechanisms needed to be complemented by distinct approaches. As we have shown, Jaspers argued that the very nature of the subject matter of psychiatry called for an understanding of the meanings of and meaningful connections between subjects' experiences. Windelband argued that understanding the values that we place in and are held by individuals required a different approach from general nomological or nomothetic science.

As with any other area of research, their suggestions were not without problems, and there has been genuine progress in developing their key ideas. Jaspers' claim that key aspects of psychopathology are both genuinely mental phenomena and yet brutally 'un-understandable' has prompted attempts to explain simultaneously how there can be at least shifting and partial understanding of still difficult phenomena. We have mentioned work drawing on a phenomenological tradition, but equally the US psychologist Brendan Maher's suggestion that delusions are an understandable response to abnormal experiences[54] can be seen as a reaction to Jaspers' work. Similarly, Windelband's idea that value judgements require a special kind of individualistic judgement has been replaced by an approach, values-based practice, that is derived from analytical philosophy and relies on 'good process', in particular learnable clinical skills, as a basis for balanced decision-making where complex and conflicting values are in play.

Our aim in giving this brief outline of the work of Jaspers and Windelband, as two key figures in the history and philosophy of science, together with their modern counterparts, has been to indicate the extent to which resources derived from philosophy can contribute alongside the sciences to improving mental health care. Phenomenology and values-based practice, furthermore, it is important to add, are themselves set within a new and rapidly expanding international field of cross-disciplinary work between philosophy and psychiatry.[21] As Box 3.3 shows, this is a rapidly expanding field, and there are other potentially important practical developments, notably in relation to tacit knowledge (the basis of professional skills) and individual judgement (as in clinical judgement).[55]

It might seem surprising to some that there should have been this expansion of philosophy and psychiatry over the past two decades, in parallel with dramatic developments in the neurosciences underpinning psychiatry. The history of ideas outlined in the first two sections of this chapter not only makes sense of these parallel developments but also helps us to see their likely future direction. Then, as now, there were dramatic (actual or anticipated) advances in the sciences underpinning psychiatry; and these advances, far from reducing the need for careful conceptual work alongside empirical studies, actually increased it. There are several reasons for this: the challenge of applying generalized scientific knowledge to particular individual human beings, but also theoretical challenges

Box 3.3 Developments in the new philosophy of psychiatry

- 43 New academic and research groups around the world
- Special sections in the WPA and AEP
- Establishment of the International Network for Philosophy and Psychiatry (INPP), launched Cape Town, South Africa, 2002
- Annual international conferences in different parts of the world
- New professorial chairs in Italy, Netherlands, South Africa and the UK
- Training and research programmes, including a recently launched Oxford DPhil
- The international journal Philosophy, Psychiatry, and Psychology (PPP) now in its fourteenth year, from Johns Hopkins University Press
- Several book series, including International Perspectives in Philosophy and Psychiatry (IPPP) from Oxford University Press
- Establishment of the Institute for Philosophy, Diversity and Mental Health (IPDMH) at the University of Central Lancashire in the UK, with over £1 million funding
- Philosophy into practice, e.g. values-based practice (see text).

within the sciences themselves – we noted at the start of this chapter Nancy Andreasen's claim that advances in the neurosciences have driven many of the deepest problems of traditional philosophy to the top of our practical agenda.

Jaspers' work in the early twentieth century represented one clear response to these challenges. Naive models of empirical science, he believed, would reduce the sciences of the day to mere scientism (you will recall that he called such models 'brain mythologies'); and his response was to argue that we needed to find ways of working with meanings alongside and in parallel with scientific work on explanatory causes. Windelband's work on values represents a different response to scientism – one built on values in so far as these are distinct from meanings. What is shared by these approaches, then, is a recognition of the need to bring together in one way or another the findings of generalized science with the uniqueness of individual human beings.

The need to reconcile generalized science with individual human beings is perhaps particularly acute at the present time, with growing pressure to base service delivery on a narrow model of evidence-based practice that Jaspers would perhaps have characterized as scientistic. Yet, again as during psychiatry's first biological phase, the dangers of a scientistic use of evidence-based practice have been most evident to those who have been at the very forefront of developing this approach. As David Sackett puts it, in his seminal training manual on evidence-based practice, it is only when best research evidence is combined with clinical skills and, importantly, patient values that 'clinicians and patients form a diagnostic and therapeutic alliance which optimises clinical outcomes and quality of life'.[56] There is no better statement of the need for combining rigorous empirical methods with equally rigorous philosophical approaches in developing a psychiatry for the twenty-first century that is both fully science-based and genuinely patient-centred.

KEY POINTS

- Jaspers divides symptoms between subjective and objective. Another person's objective symptoms can either be detected by one's senses (such as by sight or hearing) or by merely rational understanding. To be sensitive to another's subjective symptoms requires that one imaginatively puts oneself into their predicament in a way that goes beyond merely rational understanding. The distinction between objective and subjective corresponds to a distinction between explanation and understanding.
- Jaspers distinguishes between two forms of (subjective) understanding. Static understanding, also called phenomenology, concerns what mental states feel like. Genetic understanding, also called empathy, concerns the way one mental state, ideally and typically, arises from another.
- Jaspers believes that some forms of apparently mental phenomena nevertheless cannot be understood. Primary delusions are, he claims, 'un-understandable'.
- Windelband distinguishes between idiographic and nomothetic sciences. Nomothetic sciences concern law-like and general phenomena. Idiographic sciences concern one-off events or processes. The most obvious need for an idiographic approach, according to Windelband, concerns judgements about values.
- Values-based practice is a new skills-based approach to working with complex and conflicting values that has been developed from philosophical value theory to stand alongside evidence-based practice as a support-tool for clinical decision-making in mental health.

WEBSITES AND FURTHER READING

Fulford and others have examined the links between values-based practice and a number of specific areas of medicine including Child and Adolescent Mental Health Services (CAMHS), management, spirituality and other aspects of the medical humanities, ethics, diagnosis and

neuroscience. See Warwick Medical School website http://www2.warwick.ac.uk/fac/med/study/cpd/subject_index/pemh/vbp_introduction/.

Teaching and learning materials on values-based practice have now been published in a wide range of professional journals and textbooks. See the Warwick Medical School website noted above, and also the Royal College of General Practitioners' 2005 *Curriculum Statement: Ethics and Values Based Medicine* at www.rcgp.org.uk/gpcurriculum/pdfs/ethicsAndVBPsfRCGPCouncilDec2005.pdf.

Details of Jaspers' work and of other important figures in the early development of psychiatry are given in 'History of ideas' in *The Oxford Textbook of Philosophy and Psychiatry* (Fulford KWM, Thornton T, Graham G (2006) Oxford: Oxford University Press). Section 3 of this same book covers modern developments in the philosophy of science, Section 4 covers values-based practice and ethics, and Section 5 covers phenomenology and the philosophy of mind. *The Oxford Textbook* includes a series of readings (on a CD-ROM) and includes detailed guides to further reading and key learning points.

REFERENCES

1. Andreasen NC (2001) *Brave New Brain: Conquering Mental Illness in the Era of the Genome*. Oxford: Oxford University Press.
2. Kendler K, Appelbaum P, Bell C, *et al.* (2008) Issues for DSM-V: DSM-V should include a conceptual issues work group. *American Journal of Psychiatry* **165**: 1–2.
3. Department of Health (2005) *New Ways of Working for Psychiatrists: Enhancing Effective, Person-Centred Services through New Ways of Working in Multidisciplinary and Multi-Agency Contexts. (Final Report 'But Not the End of the Story'.)* London: Department of Health.
4. Department of Health (2008) *Putting People First: A Shared Vision and Commitment to the Transformation of Adult Social Care*. London: Department of Health.
5. Jaspers K ([1913] 1997) *General Psychopathology*. Baltimore, MD: Johns Hopkins University Press.
6. Ehrlich E, Ehrlich L, Pepper GB (eds) (1986) *Karl Jaspers: Basic Philosophical Writings*. Athens, OH: Ohio University Press.
7. Jaspers K ([1912] 1968) The phenomenological approach in psychopathology. *British Journal of Psychiatry* **114**: 1313–23.
8. Jaspers K ([1913] 1974) Causal and 'meaningful' connections between life history and psychosis (transl. J.Hoenig.) In SR Hirsch, M Shepherd (eds) *Themes and Variations in European Psychiatry*. Bristol: Wright: 80–93.
9. Fulford KWM, Thornton T, Graham G (2006) *Oxford Textbook of Philosophy and Psychiatry*. Oxford: Oxford University Press.
10. Carruthers P, Smith PK (eds) (1996) *Theories of Theories of Mind*. Cambridge: Cambridge University Press.
11. Heal J (1995) Replication and functionalism. In M Davies, T Stone (eds) *Folk Psychology*. Oxford: Blackwell.
12. McDowell J (1998) *Mind Value and Reality*. Cambridge, MA: Harvard University Press.
13. Davidson D (1980) *Essays on Actions and Events*. Oxford: Oxford University Press.
14. Sims, A. (1988) *Symptoms in the Mind: An Introduction to Descriptive Psychopathology*. London: Bailliere Tindall.
15. IDGA Workgroup, WPA (2003) IGDA 8: idiographic (personalised) diagnostic formulation. *British Journal of Psychiatry* **18** (suppl. 45): 55–7
16. Windelband W (1980) History and natural science. *History and Theory and Psychology* **19**: 169–85.
17. Lamiell JT (1997) Individuals and the differences between them. In R Hogan, J Johnson, S Briggs (eds) *Handbook of Personality Psychology*. New York: Academic Press.
18. Rickert H (1907) Geschichtsphilosophie. In W Windelband (ed.) *Die Philosophie im Beginn des zwanzigsten Jahrhunderts: Festschrift für Kuno Fischer*. Heidelberg: Carl Weinter.
19. Thornton T (2008) Does understanding individuals require idiographic judgement? *European Archives of Psychiatry and Clinical Neuroscience* **258** (suppl. 5): 104–9.
20. Thornton T (2008) Should comprehensive diagnosis include idiographic understanding? *Medicine, Healthcare and Philosophy* **11**: 293–302.
21. Fulford KWM, Morris KJ, Sadler JZ, Stanghellini G (2003) Past improbable, future possible: the renaissance in philosophy and psychiatry. In KWM Fulford, KJ Morris, JZ Sadler, G Stanghellini (eds) *Nature and Narrative: An Introduction to the New Philosophy of Psychiatry*. Oxford: Oxford University Press.
22. Hoerl C (2001) Introduction: understanding, explaining, and intersubjectivity in schizophrenia. *Philosophy, Psychiatry, and Psychology* **8**.
23. Widdershoven G, Widdershoven-Heerding I (2003) Understanding dementia: a hermeneutic perspective. In KWM Fulford, KJ Morris, JZ Sadler, G Stanghellini (eds) *Nature and Narrative: An Introduction to the New Philosophy of Psychiatry*. Oxford: Oxford University Press.
24. Sabat SR (2001) *The Experience of Alzheimer's Disease: Life Through a Tangled Veil*. Oxford: Blackwell.
25. Morris KJ (2003) The phenomenology of body dysmorphic disorder: a Sartrean analysis. In KWM Fulford, KJ Morris, JZ Sadler, G Stanghellini (eds) *Nature and Narrative: An Introduction to the New Philosophy of Psychiatry*. Oxford: Oxford University Press.
26. Stanghellini G (2004) *Disembodied Spirits and Deanimated Bodies*. Oxford: Oxford University Press.

27. Brown MM, Brown GC, Sharma S (2005) *Evidence-Based to Value-Based Medicine*. Chicago, IL: American Medical Association Press.
28. Hunink M, Glasziou P, Siegel J, *et al*. (2001) *Decision Making in Health and Medicine: Integrating Evidence and Values*. Cambridge: Cambridge University Press.
29. Fulford KWM (2004) Ten principles of values-based medicine. In J Radden (ed.) *The Philosophy of Psychiatry: A Companion*. New York: Oxford University Press.
30. Hare RM (1952) *The Language of Morals*. Oxford: Oxford University Press.
31. Hare RM (1963) *Freedom and Reason*. Oxford: Oxford University Press.
32. Hare RM (1963) Descriptivism. *Proceedings of the British Academy*, **49**: 115–34. Reprinted in Hare RM (1972) *Essays on the Moral Concepts*. London: Macmillan.
33. Warnock GJ (1971) *The Object of Morality*. London: Methuen.
34. Warnock GJ (1967) *Contemporary Moral Philosophy*. London: Macmillan.
35. Hume D (1972) *A Treatise of Human Nature: Book III*. London: Fontana/Collins.
36. Putnam H (2002) *The Collapse of the Fact/Value Dichotomy and other Essays*. Cambridge, MA: Harvard University Press.
37. Fulford, KWM (1989, reprinted 1995 and 1999) *Moral Theory and Medical Practice*. Cambridge: Cambridge University Press.
38. Fulford KWM (1998) Dissent and dissensus: the limits of consensus formation in psychiatry. In HAMJ ten Have, H-M Sass (eds) *Consensus Formation in Health Care Ethics*. Dordrecht: Kluwer.
39. Fulford KWM (1999) Nine variations and a coda on the theme of an evolutionary definition of dysfunction. *Journal of Abnormal Psychology* **108**: 412–20.
40. Fulford KWM (2000) Teleology without tears: naturalism, neo-naturalism and evaluationism in the analysis of function statements in biology (and a bet on the twenty-first century). *Philosophy, Psychiatry, and Psychology* **7**: 77–94.
41. Colombo A, Bendelow G, Fulford KWM, Williams S (2003) Evaluating the influence of implicit models of mental disorder on processes of shared decision making within community-based multi-disciplinary teams. *Social Science and Medicine* **56**: 1557–70.
42. Fulford KWM, Colombo A (2004) Six models of mental disorder: a study combining linguistic-analytic and empirical methods. *Philosophy, Psychiatry, and Psychology* **11**: 129–44.
43. Mezzich JE (2002) Comprehensive diagnosis: a conceptual basis for future diagnostic systems. *Psychopathology* **35**: 162–5.
44. Mezzich JE, Salloum IM (2007) Towards innovative international classification and diagnostic systems: ICD-11 and person-centered integrative diagnosis. *Acta Psychiatrica Scandinavica*, **116**: 1–5.
45. National Institute for Mental Health in England (2004) *The National Framework of Values for Mental Health*. London: National Institute for Mental Health in England.
46. Department of Health (2004) *The Ten Essential Shared Capabilities: A Framework for the Whole of the Mental Health Workforce*. London: Sainsbury Centre for Mental Health, National Health Service University (NHSU) and National Institute for Mental Health England (NIMHE).
47. Department of Health (2005) *Delivering Race Equality in Mental Health Care: An Action Plan for Reform Inside and Outside Services*. London: Department of Health.
48. Department of Health (2004) *Mental Health Policy Implementation Guide: Community Development Workers for Black and Minority Ethnic Communities*. London: Department of Health.
49. Woodbridge K, Fulford KWM (2004) *Whose Values? A Workbook for Values-Based Practice in Mental Health Care*. London: Sainsbury Centre for Mental Health.
50. Care Services Improvement Partnership (CSIP), National Institute for Mental Health in England (NIMHE) (2008) *Workbook to Support Implementation of the Mental Health Act 1983 as Amended by the Mental Health Act 2007*. London: Department of Health.
51. Van Staden CW, Fulford KWM (2007) Hypotheses, neuroscience and real persons: the theme of the 10th International Conference on Philosophy, Psychiatry and Psychology. *South African Journal of Psychiatry* **13**: 68–71.
52. Sadler JZ (ed.) (2002) *Descriptions and Prescriptions: Values, Mental Disorders, and the DSMs*. Baltimore, MD: Johns Hopkins University Press.
53. National Institute for Mental Health in England (NIMHE), Care Services Improvement Partnership (CSIP) (2008) *Three Keys to a Shared Approach in Mental Health Assessment*. London: Department of Health.
54. Maher BA (1999) Anomalous experience in everyday life: its significance for psychopathology. *Monist* **82**: 547–70.
55. Thornton T (2007) *Essential Philosophy of Psychiatry*. Oxford: Oxford University Press.
56. Sackett DL, Straus SE, Scott Richardson W, Rosenberg W, Haynes RB (2000) *Evidence-Based Medicine: How to Practice and Teach EBM*, 2nd edn. Edinburgh: Churchill Livingstone.

Chapter 4

Research methods and statistics

Daniel Stahl and Morven Leese

INTRODUCTION

In this chapter we review study designs and statistical techniques commonly used in psychiatry, focusing on the requirements of the MRCPsych examination. For reasons of space, we have not included details of all tests and definitions, but where appropriate we have given one or two key references. General texts that would be very useful to consult are: Bland (2000),[1] for a comprehensive guide to methods; Petrie and Sabin,[2] an excellent handbook covering all the standard statistical tests, including examples; and Everitt,[3] for concise definitions. Some useful articles can be found in the journals *Bandolier* and *BMJ Teaching*.

Psychiatric data: a typical scenario

We introduce here a typical scenario that will be used as an example throughout the rest of the chapter. A medication training package is offered to nurses, who each care for a caseload of patients. Each nurse willing to participate in a trial of this package is randomized to receive the training or be placed on a waiting list to receive training later, after the trial. Two patients are chosen at random from each nurse's caseload for assessment of patient outcomes. Outcomes are at two levels: a knowledge test for nurses, before and after the trial has taken place, and a global assessment of functioning (GAF) for patients. This is fairly typical of psychiatric data because it involves comparing groups at different levels within the same study, which leads to 'clustering' of data (because patients of a particular nurse are likely to be more similar to each other than patients of different nurses).

We shall use output from the package Stata to illustrate the results of analysing various types of data. Another package that is widely used is SPSS. Both programs are very comprehensive, but Stata has particularly good facilities for the types of data encountered in psychiatry. A good introduction to statistics using Stata is Dupont.[4] First we discuss some aspects of study design and general concepts such as precision and bias.

Classification of data types

The type of data being analysed determines which statistical methods are appropriate. In Table 4.1 we have given some examples of types of data that may be encountered in psychiatry and how they are classified.

Apart from the clustering feature mentioned above, psychiatric data have other characteristics that make them somewhat distinct from medical data in general. One of

Table 4.1 Data types encountered in psychiatry

Type of scale	Example	Data	Relation between values
Nominal	Gender, diagnostic group	Discrete (categorical, including binary)	No particular order; values are same/different
Ordinal	Age group, Likert scale, adjectival scale	Discrete (ordered categorical)	Categories are ranked – (ordered data); values are smaller/larger
Interval	Visual analogue pain score, functioning score	Discrete or continuous	Absolute difference; 0 has no real meaning
Ratio	Number of symptoms, cost of treatment	Discrete or continuous	Absolute difference and proportion; 0 has real meaning

these is the widespread assessment of patient outcomes from *questionnaires*. Establishing validity and reliability is particularly important for such data, as are issues concerned with combining responses to several individual questions (or *items*) into an aggregate score. Before describing the design and analysis of collected data, we first consider some aspects of scale development.

SCALE DEVELOPMENT

Measurement scales

Measurement involves the estimation of a quantity relative to some standard. Measurements usually need a *measuring instrument*, such as a measuring tape. The type of unit on which something is measured is called the *scale* (e.g. metre, second, kilogram) of the measuring instrument. In psychiatry, typical measurement problems involve identifying and assessing the severity of a disorder and evaluating the outcome of a treatment. However, many traits, such as abilities, attitudes, quality of life and personality, are difficult to measure, since they are unobservable and subjective. Psychometrics is concerned with the theory and technique of measuring such psychological phenomena, which are called *constructs*, *latent traits* or *attributes*. In psychometrics, unobservable and subjective phenomena are *inferred* from variables (items) that can be observed and measured, in the case of psychiatry through the use of questionnaires and structured interviews. Such questionnaires are sometimes called *instruments*. Adjectival, Likert and visual analogue scales (VAS) are often used in such situations. Figure 4.1 shows an example of each of these.

An example of a published adjectival scale is the Beck Depression Inventory (BDI).[5] This consists of a 21-question multiple-choice self-report inventory and asks questions about the feelings of the subject during the past couple of days. Each question (item) offers four possible answers, arranged in increasing severity. An example of statements for one item is:

0 – I do not feel like a failure.
1 – I feel I have failed more than the average person.
2 – As I look back on my life, all I can see is a lot of failure.
3 – I feel I am a complete failure as a person.

In Likert scales, a declarative sentence is given along with a number of response options. Respondents specify their level of agreement to this. While the descriptor of the adjectival scale is unipolar (from low to high), a Likert item is bipolar, with a neutral/undecided/no opinion term in the middle.

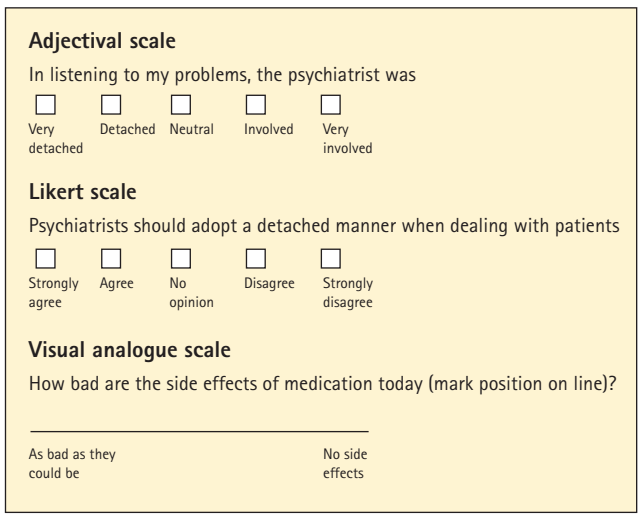

Figure 4.1 Examples of questionnaire data commonly used in psychiatry

Constructing a measuring instrument

The aim is to develop a scale that is *reliable* (i.e. measuring consistently) and *valid* (i.e. measuring what it is supposed to measure). The first step in the development of a questionnaire is to generate a pool of items, which is assumed to measure the latent construct(s).[6] From a statistical point of view, the choice of the scale (e.g. adjectival or Likert scale), the level (continuous, ordinal or categorical), the definition of the scaling responses (e.g. choice of adjective or the number of answer steps) and the wording of the question will influence the quality of the final scale. Particularly troublesome are floor and ceiling effects in short scales, where respondents tend to use a compressed interval at the top or bottom of the scale; this limits variability and hence the discriminatory power of the questionnaire.

To investigate how items might relate to latent constructs, it is common to perform a factor analysis, either exploratory (if little is known about how the items relate to factors) or confirmatory (if theory can predict this). Once the items that measure the construct(s) have been decided on, it is then common practice to add them up to make a scale total and/or subscale totals. In calculating totals, it is important that the response values (e.g. from 0 to 5) are in the same conceptual order to obtain a meaningful total score, so some items may need to be *reverse-coded* (e.g. 5 to 1, 4 to 2, 3 to 3, 2 to 4, 1 to 5). A summary of issues relating to health measurement scales in general is given in Streiner and Norman.[7]

Types of reliability

A reliable measure is measuring something consistently, in a reproducible fashion. In psychiatry, many questionnaires are self-administered and the reliability of the respondent needs to be evaluated by administering the scale on two or more occasions. This is called *test–retest reliability*. The choice of the time interval between two administrations is crucial. It must be sufficiently long that respondents do not remember their previous answers, but short enough that the

underlying process (such as degree of depression) has not changed. Typical test–retest time intervals for health-related studies range between 2 days and 14 days. Often tests are administered through structured interviews. In this case, one needs to assess the degree to which different raters give consistent estimates of the same phenomenon when rating the same patients (*inter-rater reliability*). Finally, the individual items need to reliably estimate the latent construct. *Internal consistency reliability* assesses the consistency of items within a scale or subscale.

Indices of reliability

Classical test score theory assumes that each person has a true score that would be obtained if there were no errors in measurement. However, measuring instruments are imperfect and an observed score will be the sum of the true score and error. *Reliability* reflects the relative contributions of those two sources of variation. Reliability for continuous variables is usually based on the *intraclass correlation coefficient* (ICC).[8] It is the proportion of variance that reflects between-subject ('true') variability and ranges between 0 (no reliability) to 1 (perfect reliability). The most commonly used measure is the ICC for absolute agreement, which measures the absolute agreement between raters or between two observations at different time points and is the between-subject variance divided by the total variance. If we are interested only in relative agreement (i.e. ignoring constant bias between the raters or time points), then the between-rater or between-time-point variance would be removed from the model. The ICC for relative agreement would be estimated by the between-subject variance divided by the sum of between-subject variance and within-subject variance, and is consequently larger than the ICC for absolute agreement. Another statistic that can be used for pairs of observations (different occasions or two raters) is the so-called *concordance index* (see Bland and Altman[9]); this is numerically similar to the ICC for absolute agreement.

Although many software packages calculate ICCs, they can also be estimated through *random effects models* (see Multivariate and other more complex techniques, below), which can easily extend the ICC to account for more than two factors. It is, for example, possible to include therapist as another random factor in the model to estimate sources of errors due to therapist effects. This extension of reliability is known as *generalizability* theory and is a powerful tool to identify sources of error in a measurement process.[10]

The test–retest or inter-rater reliability of categorical variables is evaluated using *Cohen's kappa*, a measure of chance-adjusted agreement that takes a value of 1 when there is perfect agreement and of 0 when observed agreement is equal to chance. Ordinal categorical variables, such as symptom scores with possible scores of low, medium, high, can be assessed using *weighted kappa*, which penalizes according to the extent of disagreement. The kappa coefficient is dependent on the prevalence of the different response types. For example, if the prevalence of depression is low, we can obtain a good reliability for the diagnosis of depression between two raters, even if the raters disagree strongly about the classification of the few positive cases of depression. It is therefore advisable not to rely solely on the kappa but also to assess the percentage agreement within each class.

For internal consistency, a common approach is to calculate Cronbach's alpha for continuous data, which is the average of all possible split-half reliabilities (where the items are divided in half, e.g. odd versus even questions, and the reliabilities computed for the two halves). If the data are binary, the formula is the same but the index is known as the *Kuder–Richardson Formula 20* (KR20). Because reliability increases with the number of items, the index can be adjusted by the *Spearman–Brown formula*, so that the reliability of scales with different lengths can be compared directly.

Table 4.2 summarizes the various types of reliability discussed.

Intraclass correlations can be classified as 0.6 (fair), 0.7 (good), 0.8 (very good) and 0.9 (excellent), although note that these are all subjective terms; kappas and Cronbach's alphas are sometimes classified in a similar way.

Types of validity

Reliability does not say anything about what is being measured. A measure also requires validity – that is, finding

Table 4.2 Summary of reliability measures

Reliability	What it measures	Typical indices
Inter-observer	Agreement among interviewers, raters or coders who are rating the same information	Intraclass correlation (continuous data)
Test–retest (or intra-observer)	A measure at two different times with no treatment (or other changes) in between should yield the same results	Kappa (categorical data; possibly weighted if more than two categories)
Internal consistency (or item reliability)	Do the items measure the same construct/domain? Alternatively, do the scores of the items correlate with scores of all other items of the same construct/domain?	Cronbach's alpha (or Kuder–Richardson for binary data); possibly corrected using the Spearman–Brown formula

results that accurately reflect the concept being measured. Reliability is irrelevant without validity and validity implies reliability. Thus, reliability places the upper limit on validity. Some of the types of validity are summarized in Table 4.3.

The evaluation of validity is more of a scientific issue than a statistical one, since it depends on choosing appropriate scales for comparison. It is typically assessed through correlation or regression analysis, comparing the results of the scale of interest with results from other scales or designing experiments to reveal facets of the causal role of the construct. A common validity test involves comparing a screening test against a gold standard diagnosis that has established the true condition, for example as shown in the *contingency table* or cross-tabulation in Table 4.4.

Diagnostic test measures

In a diagnostic test (and other situations where one predicts the value of a binary variable), there are a number of measures of predictive accuracy. Sensitivity (specificity) is the proportion of those with (without) the true diagnosis who screen positive (negative). The (positive) likelihood ratio (LR) is the value of the screening test in predicting a positive result and is the ratio of the probability of a positive test from someone with the disorder to the probability of a positive test from someone without the disorder (and similarly for the negative LR). These measures are independent of prevalence and so are generalizable across samples. Positive (negative) predictive power is more useful in clinical situations and is the proportion of those who screen positive (negative) who do (do not) have the diagnosis. Positive predictive value (PPV) and negative predictive value (NPV) are specific to the particular clinical situation since they depend on prevalence. Table 4.5 summarizes these various quantities.

The post-test probability is a useful measure for clinicians in assessing a positive test result for individual patients. It takes account of known prevalence and combines it with the test result, through the intermediate step of computing odds from probabilities, using the following relationships:

$$\text{Odds} = \text{probability}/(1 - \text{probability})$$
$$\text{Probability} = \text{odds}/(1 + \text{odds})$$

The relevant prevalence is that for the specific population of interest or may be obtained from similar samples. The post-test odds, and hence the post-test probability, can be calculated as follows:

$$\text{Post-test odds} = \text{pre-test odds} \times \text{likelihood ratio}$$

If a patient undergoes a series of tests, the LRs can be multiplied together to obtain the post-test probability based on all the tests simultaneously.

For example, suppose the prevalence in the population of interest is 20 per cent, so the pre-test probability of the condition is 0.2 and odds is 0.2/0.8 = 0.25. A patient tests positive using a test that has sensitivity 0.8 and specificity 0.9, and that therefore has $LR_{+ve} = 0.8/0.1 = 8$. The post-test odds

Table 4.3 Types of validity

Type of validity	What it measures
Face	Subjective assessment that the instrument/item appears to measure the desired qualities
Content	Subjective assessment that the instrument samples all the important contents/domains of the attribute
Criterion	Is the measure consistent with what we already know and what we expect?
Concurrent	Does the new measure associate with a different scale that measures something similar?
Discriminate	The new measure should not associate with constructs that should not be related
Predictive	Does the measure predict outcome accurately?
Construct	Is the measure related to other variables as required by theory? Sensitivity to change, responsiveness

Table 4.4 Example of cross-tabulated data as used in a screening test

	True positive	True negative	Total
Screen positive	a	b	$a+b$
Screen negative	c	d	$c+d$
Total	$n_1 = a+c$	$n_2 = b+d$	$a+b+c+d$

Table 4.5 Diagnostic test measures

Measure	Formula
Sensitivity	$a/(a+c)$
Specificity	$d/(b+d)$
Likelihood ratio for a positive test	Sensitivity/(1 – specificity)
Likelihood ratio of a negative test	(1 – sensitivity)/specificity
Positive predictive power	$a/(a+b)$
Negative predictive power	$a/(c+d)$

is $8 \times 0.25 = 2$. Using the relationship above, the post-test probability is therefore $2/(1 + 2) = 0.667$ (67%).

Finally, we illustrate a receiver operator curve (ROC) in Figure 4.2. This is a plot of sensitivity versus 1 – specificity for situations where the screening tool produces a continuous score. In that case, by using different cut-offs, one can balance sensitivity and specificity. The area under the curve is an overall measure of the predictive success of the tool. The example in Figure 4.2 is not particularly good at 0.609: perfection would be 1, and 0.5, the diagonal line, no better than chance. The further the curve is towards the left-hand top corner, the better (high sensitivity and low specificity).

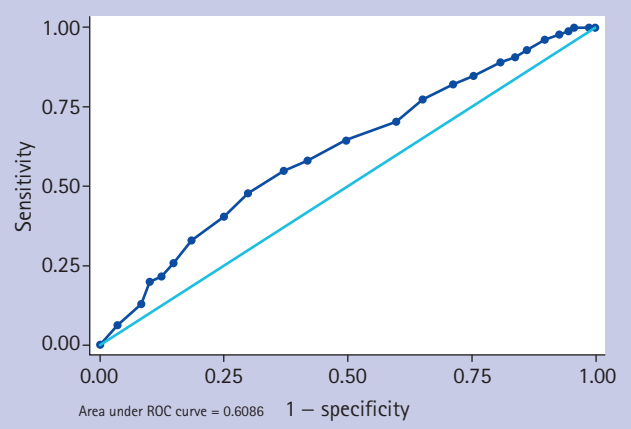

Figure 4.2 Example of a receiver operator curve (ROC)

STUDY DESIGNS

Although some psychiatric studies can be described as exploratory, most aim to answer specific questions such as: Does this new therapy work better than a standard therapy already in use? Can we identify factors that are risk or protective factors for developing certain diseases? Is it worth spending money on specific therapies or methods of delivering services? These three questions most often lead to intervention, observational and health economics studies, respectively. This classification is somewhat simplistic, as there are often going to be overlaps, but it provides a convenient way of thinking about study design. See Pocock[11] for an introduction to clinical trials.

Intervention studies

The main types of intervention (or 'experimental') study are now described.

In a *randomized controlled trial* (RCT), random allocation of patients to the treatments to be compared ensures fair comparisons. Any differences will be due to chance, and with large samples they will tend to be small. A *crossover* design involves giving the intervention and control to two randomized groups as usual but then, after a washout period, patients swap over, so that the controls have the intervention and vice versa. The advantage is that the difference between intervention and control is measured for each patient, so that patient differences are controlled for in the design. It is only possible for certain types of intervention (mainly drugs whose effect will become negligible during the washout period).

Randomization may be performed separately within strata (e.g. by centre in a multicentre trial, or by gender and/or age group). Methods of randomization include *permuted blocks* and *minimization*. The former produces sequences of codes (e.g. ABBA) that are generated at random and used to allocate patients to treatment arms A and B in small blocks, in this case four. Minimization is not a truly random method, but nevertheless it is generally acceptable and allows easy balance across several strata simultaneously. *Double blinding* means that neither the patient nor the therapist knows the group to which the patient has been assigned; a further protection is triple blinding, where the person analysing the data is also blind. The double-blind randomized, controlled trial is the gold standard of experimental designs since it minimizes most types of bias.

The interventions to be compared may include a *placebo* – that is, a dummy treatment that appears similar to patients but that has no medical effect. In *open-label studies*, both the patient and the therapist know the group to which the patient has been assigned. An example of the latter might be a comparison of cognitive-behaviour therapy (CBT) with drug treatment for anxiety; clearly neither the patient nor the therapist will be blind to the treatment arm and in this case a placebo is not feasible.

An ethical principal for RCTs is that there should be *equipoise* – a genuine uncertainty in the minds of the doctors concerned (whether specialists or the wider scientific community) about which arm would most benefit a subject. Another key issue is *intention-to-treat* (ITT): this refers to the recommended practice of analysing subjects as randomized, whether or not they are receiving the allocated treatment. This means that the benefits of randomization are retained, but it also means that it is the offer of treatment, rather than the receipt of treatment, that is being tested.

Trials are classified into phases depending on the stage of development of the intervention. The classification shown in Table 4.6 is most often used for drug trials. See also the Medical Research Council (MRC) Clinical Trial Unit definitions at www.ctu.mrc.ac.uk, from which some of these definitions were taken.

The reporting of trials is covered by the CONSORT guidelines,[12,13] which include a diagram showing the numbers recruited, randomized and followed up (Figure 4.3).

Loss to follow-up has traditionally been dealt with using *last observation carried forward* (LOCR), but this is now discredited and so-called *principled* methods of dealing with missing values are recommended, such as multiple imputation, model-based methods such as maximum likelihood with predictors of being missing, and probability weighting.

Table 4.6 Classification of phases in randomized controlled trials

Phase	Aim	Comment
I	Test the safety of a new treatment	Involve only a small number of people, who may be healthy volunteers
II	Test the new treatment in a larger group of people who usually have the disease for which the treatment is to be used, to see whether the treatment is effective	Usually a few hundred people are involved at this stage; phase II trials also look at safety
III	Test the new treatment in a larger group of people	Compares the new treatment with the treatment currently in use, or occasionally with a placebo; these trials look at how well the new treatment works and at any side effects it may cause
IV	Tests drugs that are already available for prescription (rather than new drugs under development)	Done after a drug has been tested in phases I–III and has been granted a licence

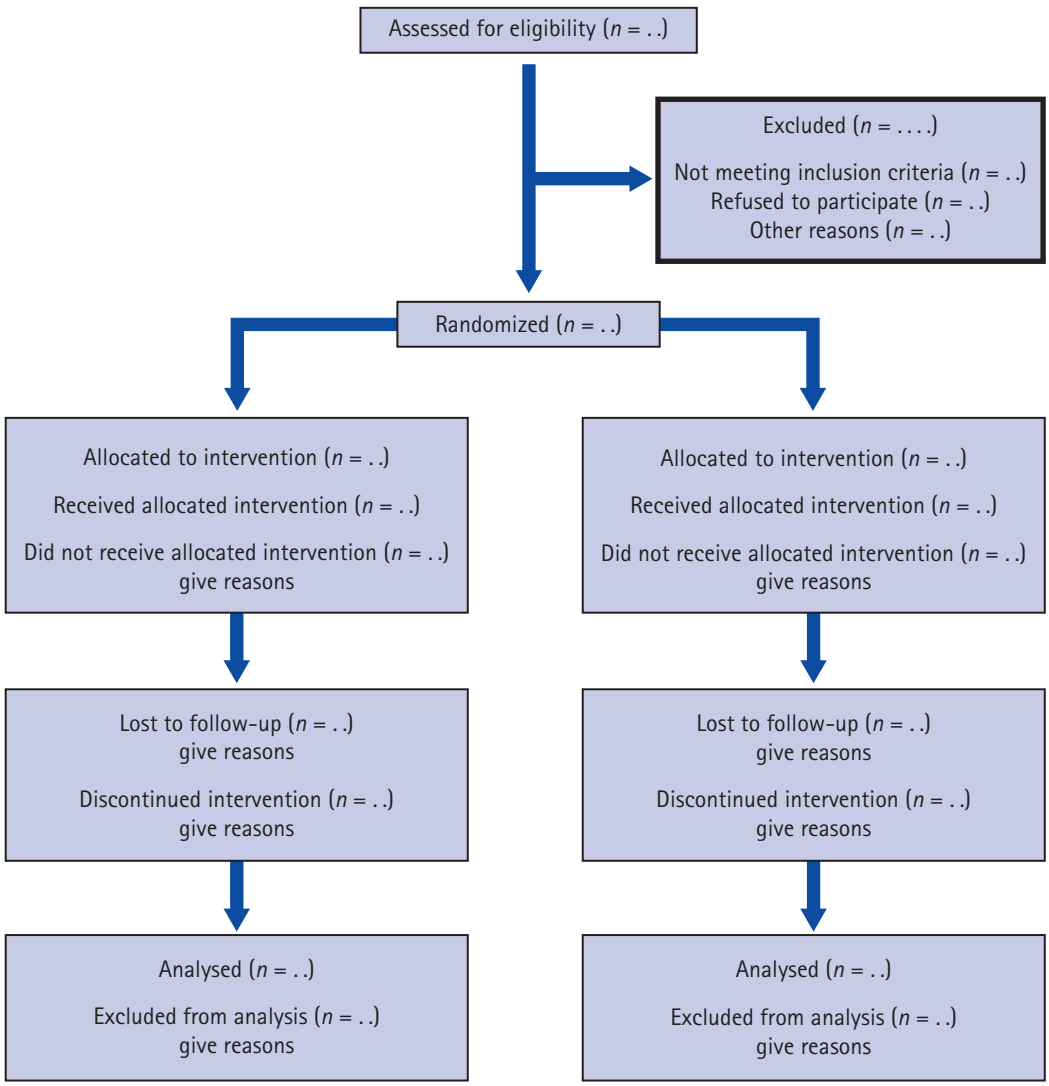

Figure 4.3 Example of a CONSORT diagram for an individually randomized trial

In a *cluster randomized trial*, a group of patients (e.g. those in a particular hospital ward) are randomized together; outcomes may be measured at the patient level (e.g. symptoms) or the group level (e.g. ward atmosphere), or both. This type of design is relatively common in psychiatry, since therapies are often administered to whole groups at a time. The phrase 'clustered data' should not be confused with 'cluster analysis' (see Multivariate and other more complex techniques, below). Although randomization can balance the characteristics of the groups across the arms of the trial, it may not be sufficient to balance those of patients. This is a disadvantage compared with individually randomized trials, although practical considerations or a need to avoid *dilution* may override this. Dilution refers to the possibility of patients randomized to the different arms interacting with one another and sharing some of the effect of therapy. A slightly different version of the CONSORT diagram would be used for clustered trials,[14] and sample sizes need to be increased to take account of the lack of independence within clusters.[15]

In *non-randomized controlled studies*, two groups are compared but individuals are not allocated at random; rather, the intervention is given to a group of patients and another group is chosen to be as similar in all respects to the intervention group as possible. Data may already exist (*historical controls*) or may be gathered at the same time (*concurrent controls*). Sometimes this design is the only feasible option, but it may be associated with bias (see Deeks *et al.*[16] for a discussion).

Patient preference trials allow patients or doctors to choose the arm of the trial according to their preference; clearly this design loses the bias-reduction advantages of randomization, and it tends to be used only where randomization is unethical or would lead to impossibly low recruitment rates. Sometimes preference trials are combined with RCTs to give four groups: those who are content to be randomized (two groups) and those who choose for themselves (two groups).

Before–after (pre–post) studies measure an outcome on the same group of patients before and after an intervention. Patients are sometimes described as being 'their own controls'. Major disadvantages of this approach are that one cannot distinguish the effect of the intervention from (i) natural improvement over time and (ii) *regression to the mean*. The latter phenomenon occurs where improvement of some patients is inevitable because their initial symptoms were high by chance. See Morton and Torgerson[17] for a discussion of this phenomenon.

Observational studies

Observational or epidemiological designs are used to describe populations, to investigate risk or protective factors, or to study interventions when it is not feasible or ethical to perform an RCT.[18]

In *case–control studies*, a group of cases is identified and then a comparison group of controls is assembled; if each case is matched individually with one or more controls, this is known as an *individually matched* case–control study. The important point is that a control would be a case *if* they had the outcome of interest. By comparing two groups with respect to hypothesized risk factors, one can infer something about the aetiology of the disease. If the control group is similar in general respects to the cases (e.g. they may be chosen to have a similar age range and to live in similar areas) without being individually matched, this is sometimes known as *group matching*. The analytical methods for an individually matched case–control study would be to estimate odds ratios (ORs) for the risk factor in relation to case–control status using, for example, conditional logistic regression (see Correlation and regression, below). Positive ORs indicate increased risk for those with the factor, but note that ORs can be numerically identified with risk ratios (RRs) only when the risk is small, say less than 5 per cent.

A typical scenario might be to identify cases of schizophrenia and find an individually matched control for each of them of the same age, within ± 2 years, of the same ethnic group and born in the same postcode area. One might then look at birth records to determine whether birth trauma had occurred. Higher odds of birth trauma among those with schizophrenia compared with controls might imply that this was a risk factor.

Retrospective cohort studies compare outcomes in a group of people who have been exposed to a risk factor and another group who have not. An example would be a study of students in which previous neurotic symptoms in adolescence are self-reported. Current diagnosis of schizophrenia is made by a psychiatrist, and the association between the earlier self-reported symptoms and schizophrenia is estimated. *Prospective* cohort studies look forwards rather than backwards in time and collect data on patients as they become exposed to the risk factor, comparing the outcomes after the passage of time. An example might be a study of people from retirement comparing time to death for those with diagnosis of Alzheimer's disease with those without. Prospective studies are preferable scientifically because there is less danger of recall bias (see below), but they may be impractical because of the need to wait until data on sufficient people have been collected. Figure 4.4 contrasts the designs of case–control and cohort studies.

Cross-sectional studies take a snapshot in time in order to investigate associations between risk factors and outcomes or to estimate the prevalence of a condition. For example, a survey might seek to detect all people in contact with mental health services in a specific area over a 6-month period; given the total population figures for the area, this could yield an estimate of the *period prevalence* of severe mental illness. If health service providers were asked how many new cases of severe mental illness occurred in the period, this would yield an *incidence rate* per year. If patients' sociodemographic details were recorded, the

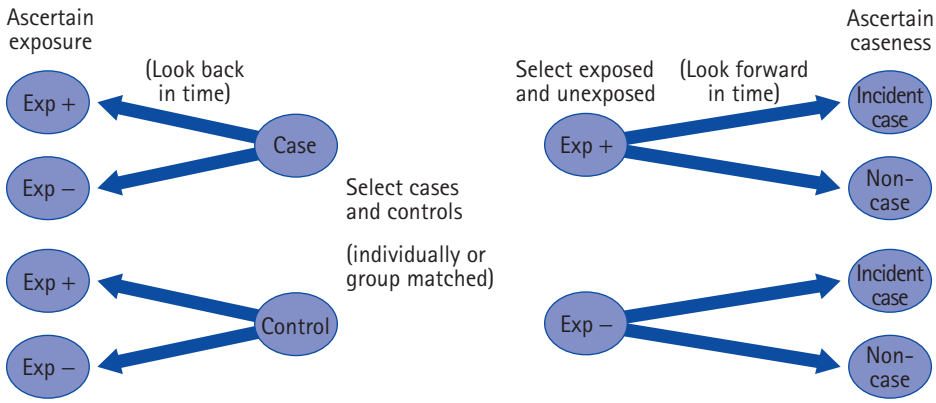

Figure 4.4 Contrast between the design of case–control (left) and cohort studies (right)

association between ethnic group and particular types of mental illness could be estimated (an *association* study).

Where the unit of observation is an area or group of people, such as a general practitioner's (GP) practice, rather than an individual, the study is termed *ecological*. Such studies are most useful for health service provision, where conclusions are sought at institutional level so that changes might be implemented at that level. The *ecological fallacy* refers to the incorrect attribution to individuals of findings at the group level. For example, the link between the levels of presentation with depression among adolescents might be studied at health authority level, with local information campaigns as a possible predictor. If a link were found, one could increase campaigns as a public health measure. The fallacy would be to assume that information campaigns affected individual adolescents when in fact their influence was more diffuse. Confounding is a particular problem for ecological studies, although they are convenient to perform because routinely collected administrative data can be used.

Health economic studies

Health economic studies are usually focused on comparing the costs and consequences of competing courses of action, such as the use of a new drug, or ways of providing health services, such as the use of assertive outreach teams. See Phelps[19] for a general introduction.

The consequences in a *cost-effectiveness analysis* might be, for example symptom reduction or social functioning. Cost-effectiveness analysis thus tends to be performed where there is a specific disease-related outcome. For example, if a new drug to lessen symptoms in schizophrenia were tested, the effectiveness might be the reduction in a symptom score. Typically, cost-effectiveness is measured as the average extra cost per unit increase in effectiveness – the incremental cost-effectiveness ratio (ICER). If this is less than a *maximum willingness to pay*, the therapy is considered cost-effective at that level. *Cost-utility analysis* is similar to cost-effectiveness analysis, except that, in order to place treatments for different diseases on a common basis, outcomes are measured in generalized units of utility. A commonly used utility is the *quality-adjusted life year* (QALY). One QALY is equivalent to 1 year of perfect health or 2 years in a health state valued at 0.5.

Typically, the results are displayed in a cost-effectiveness plane, as in Figure 4.5. Here, many samples consistent with the data have been *bootstrapped* to indicate the degree of uncertainty in the overall cost-effectiveness figure (the bold dot). Bootstrapping is a technique whereby many subsamples are taken from the observed data in which some cases are dropped and some replicated; each subsample supplies point estimates, the variation in which indicates the uncertainty in the data; it is used where distributional assumptions of standard statistics cannot be met – as is often the case for cost data. Two lines showing different values for willingness to pay are also shown. Points below these lines would be cost-effective using these criteria.

In *cost–benefit analysis*, clinical consequences are measured in monetary terms and can therefore be set directly against cost. It is used where different health technologies are to be compared with each other and/or with other types of expenditure. *Cost-minimization analysis* may also be a part of economic evaluation, where the health technologies to be compared have equal benefit, so that one requires only to compare costs. An increasingly widespread technique is the use of the *cost-effectiveness acceptability curve* (CEAC), a plot of the probability of cost-effectiveness against various choices of minimum willingness to pay.

IMPRECISION, BIAS AND CONFOUNDING

Psychiatrists want to know whether the observed advantages of a new therapy in comparison with one already existing can be generalized to all patients and need to be confident that the new therapy will work (on average) for all patients with the same diagnosis. A *population* in statistics represents all units of interests, such as all patients with a similar diagnosis or all nurses in the UK who care for patients with dual diagnosis. Usually we cannot study the entire population, and we have to draw a random sample

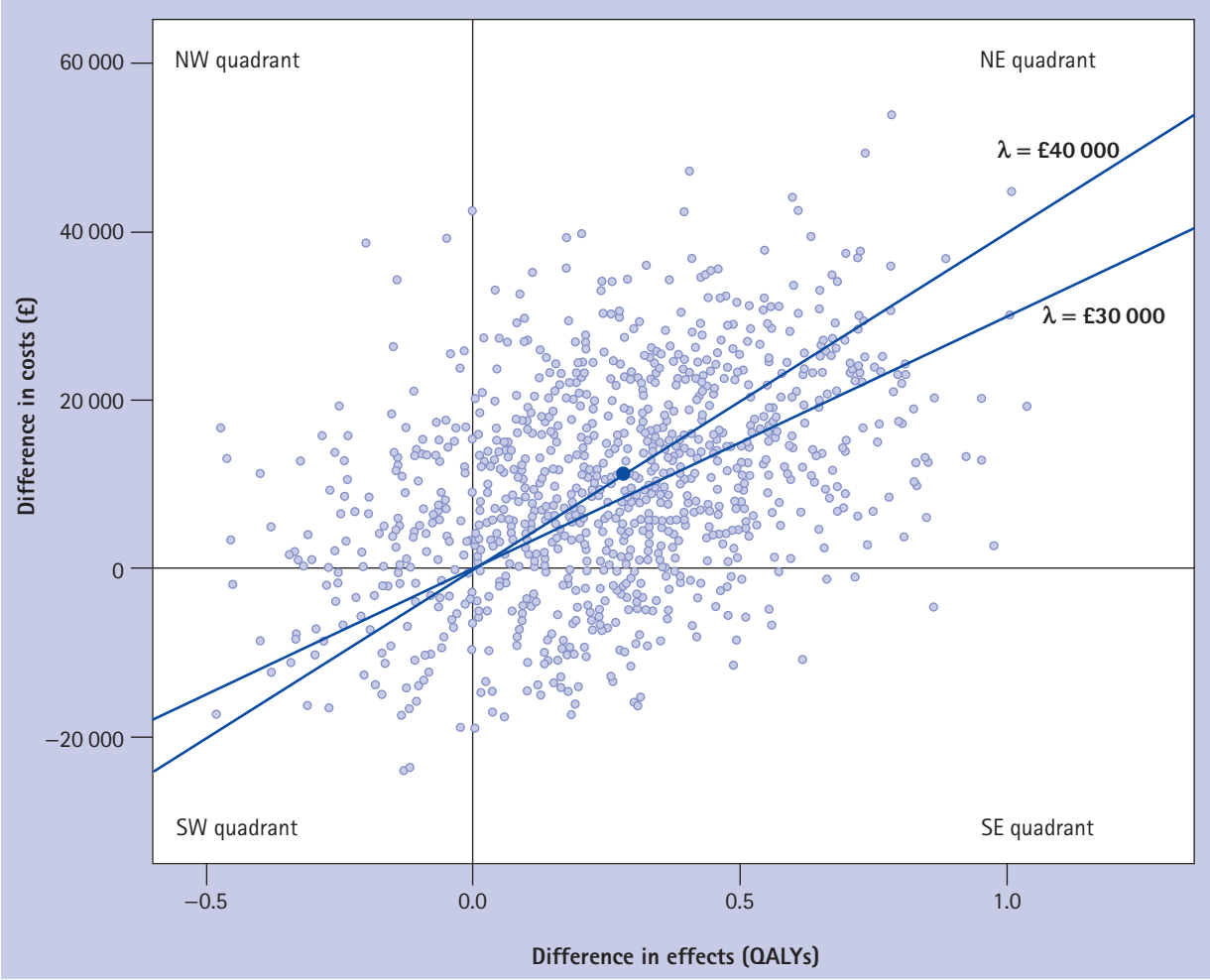

Figure 4.5 A cost-effectiveness plane showing bootstrapped samples

from a defined population, for example a sample of 30 nurses. Outcomes for the sample are assessed in relation to risk factors or interventions, and conclusions are assumed to be valid for the underlying population.

How far the conclusions can be justified depends on imprecision, bias and confounding. It is part of the job of the investigator to identify the sources of these errors and to control for them, as far as possible, either through an appropriate design or through the use of a particular type of analysis. On top of all of these factors is chance or random variation, which can affect results despite perfectly measured data, and the possibility of a conceptual error. An example of the latter might be reverse causality, when A is assumed to cause B because of an association between them, when in fact B causes A. Figure 4.6 shows a summary of these various factors.

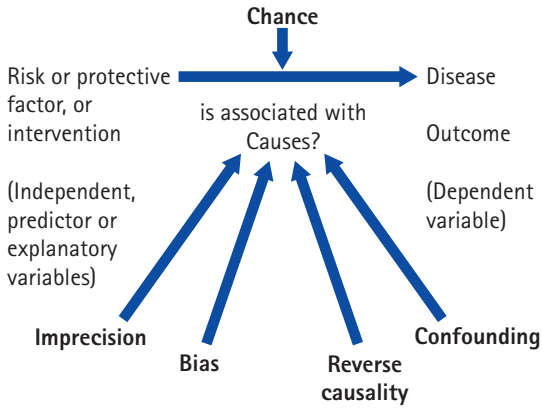

Figure 4.6 Factors that can affect inferences from data

Imprecision

Suppose we measure the mean knowledge before training for a random sample of nurses. This can be used to estimate the mean for the underlying population – that is, all nurses who care for similar patients. However, due to *sampling variation* (random fluctuations), parameter estimates differ from the true values in the population, and different studies will reveal different estimates. Precision can be increased by using larger sample sizes, since imprecision reflects the type of error that will be averaged out in summary statistics. Two

measures are commonly reported to quantify the precision of parameter estimates: standard errors and 95 per cent confidence intervals (see Statistical inference, below).

Bias

Bias – a systematic error in results or inference – is potentially more serious than imprecision, since it will not average out. Often little can be done other than to recognize that bias may have occurred and to make due allowance in interpretation. Many sources of bias have been classified and given names. Some important ones are described below: see Sackett[20] for a more complete list and Lee *et al.*[21] for information about bias in case–control studies in psychiatry.

- *Attrition bias:* the loss to follow-up of subjects from a study (once the population to whom the study applies has been defined). Reasons for attrition include illness that prevents contact, death (unless this is the primary outcome), a move to another area, the patient becoming well and being less interested in participation, and some other personal reason for unwillingness to continue. Attrition is a common problem in long cohort studies, but it can also occur in randomized controlled trials.
- *Ascertainment bias:* occurs when the disease or risk factor is not perfectly identified. This would be a problem in the Alzheimer's cohort study mentioned above if it were designed as a prospective study, because of the difficulty of diagnosing Alzheimer's except postmortem.
- *Berkson's bias (or admission rate bias):* a spurious association may be inferred because the case data arise from a special source. This type of selection bias occurs especially in case–control studies based in hospitals; for example, if in-patients with schizophrenia are matched with controls, also in hospital, in order to investigate cannabis use as a risk factor. However, cannabis use in itself may tend to lead to admissions and therefore is seen more frequently among those in hospital.
- *Information bias:* a systematic difference in the information obtained for different groups of subjects. For example, risk factor data may be either easier or more difficult to obtain for controls than cases. *Recall bias* is a type of information bias particularly associated with case–control studies based on retrospective data. For example, parents of people with schizophrenia may be more likely to attribute the illness to childhood experiences because they have been searching for an explanation; on the other hand, they may be less likely to recall unfavourable circumstances such as poor nutrition if they felt themselves to be responsible. *Observer* or *interviewer bias* is another specific example of information bias, where the information gathered is affected by the views or experience of the observer.
- *Selection bias:* when those who are selected for a study are not representative of those to whom the conclusions are to apply. If each person in the population to whom results are to apply has an equal chance of being selected, then the sample is said to be representative. This is difficult to ensure, and representativeness is often tested after data collection by comparing characteristics of the sample obtained with the wider population. An example might be a study where advertisements are placed in the press for subjects. Only the section of society who read newspapers will be recruited. *Membership bias* is a subtype of this type of bias: people who choose to be a member of a group may tend to have particular characteristics, which differ from those people with the illness in general. For example, members of user groups may derive psychological advantage from campaigning on behalf of their fellow sufferers.

Causal inference and confounding

Causality can rarely be inferred directly from association with complete confidence. Some criteria for causal inferences have been suggested by a number of authors, including the famous epidemiologist Bradford Hill.[22] Table 4.7 lists Hill's criteria (the comments are from the useful discussion by Höfler).[23]

The applicability of the above needs to be judged in relation to each study, and those in Table 4.7 are perhaps common sense. However, confounding is a key issue that is ever present, although not always obvious. *Confounders* are variables associated with the outcome of interest and also with the risk factor or intervention. For example, keep-fit classes for patients with depression might be more attractive to young people than to middle-aged or older people. If young people were more likely to recover spontaneously from depression, then one might wrongly attribute a good outcome to the keep-fit classes rather than to youthfulness. Usually a confounding factor will produce a spurious positive finding, but in some cases confounding factors can mask a real association. Randomization (in intervention studies) and matching (in epidemiological studies) can average out unknown (or known but unmeasured) confounders to some degree. Analytical techniques such as regression can also deal with remaining confounding to some extent, by controlling for them. However, *residual confounding* can still occur when a particular factor has not been controlled for (perhaps because it was never measured or recognized or because it was measured inaccurately).

Sometimes intermediate variables are associated with both outcome and the explanatory factor of interest (e.g. intervention group or risk factor) but they lie on the *causal pathway*. Such variables are not considered as confounders and would not be controlled for in the analysis or design if the aim were to assess the increased risk associated directly with the factor of interest. For example, in a study of association between antipsychotics and bone fractures in elderly people, in order to predict who might need additional interventions, one would not control for bone density in assessing the risk of antipsychotics. This would control away the effect of interest. Finally, one would usually

Table 4.7 Bradford-Hill criteria for establishing a causal relationship

Criterion	Interpretation
Strength of association	A strong association is more likely to have a causal component than is a modest association
Consistency	A relationship is observed repeatedly
Specificity	A factor influences specifically a particular outcome or population
Temporality	The factor must precede the outcome it is assumed to affect
Biological gradient	The outcome increases monotonically with increasing dose of exposure or according to a function predicted by a substantive theory
Plausibility	The observed association can be plausibly explained by substantive matter, e.g. biological explanations
Coherence	A causal conclusion should not fundamentally contradict present substantive knowledge
Experiment	Causation is more likely if evidence is based on randomized experiments
Analogy	For analogous exposures and outcomes, an effect has already been shown

assume that, in a well-thought-out study, the exposure has the potential to cause the outcome. However, sometimes it is necessary to consider reversing the outcome and the exposure (*reverse causality*). An example might be a hypothesized causal relationship between cannabis and schizophrenia: does cannabis cause schizophrenia through a direct action on the brain, or are people with mental illness more likely to take cannabis because it helps with the side effects of medication?

DISTRIBUTIONS AND SUMMARY STATISTICS

In a clinical trial, the first table presented in a report might contain estimates of the means, standard deviations, frequencies and proportions for key variables for the randomised groups at baseline. (Contrary to some practice, there should be no significance tests here, because it can be assumed that any differences, even if large, are due to random fluctuation.) The second table might contain the estimates of the mean difference between the groups at follow-up (after the intervention), its 95 per cent confidence interval and an associated *P* value. The first table provides essentially descriptive information. The second provides information from which *inferences* can be made – because in that case we hope to come to some conclusion about the effectiveness of the intervention in the population at large. This is discussed later in the section Statistical inference; here, we give some information on summary statistics and distributions that describe data.

Distributions describing data

In many datasets encountered in psychiatry, the underlying population distribution from which such data arise is the bell-shaped *normal* or *Gaussian* distribution. An example is shown in Figure 4.7. It is fundamental to statistics because it underlies many standard summaries and test statistics. In the normal distribution, most of the cases are close to the centre and relatively few examples are at the extremes. It is characterized by the mean μ (the central value or average) and the standard deviation σ, a measure of variation in the same units as the data, and is denoted $N(\mu,\sigma)$. The variance is the square of the standard deviation. Conventionally, parameter values for the population (true but usually unknown) are denoted by Greek letters and estimates from samples by lower-case roman letters.

In a normally distributed population:

- 68 per cent of the cases in the population lie within $\pm\sigma$ from μ;
- 95 per cent of the cases in the population lie within $\pm 1.96\,\sigma$ from μ.

One reason why the normal distribution is so important is the *central limit theorem*, which states that under certain conditions the sum of a large number of independent and identically distributed random variables will be distributed approximately normally. Thus, the distribution of a mean tends to be normal, even when the distribution from which the means are derived is not, and it has the same mean as that of the original population. The theorem also explains why scores that reflect a diverse set of underlying additive biological or psychological influences emerge as approximately normal.

Two other distributions that are important in describing data are the Poisson and binomial distributions. The *Poisson distribution* describes counts that arise over time, in space or as rare events in a population. Examples might be the number of hospital admissions over a person's lifetime, the number of lesions in a particular volume of the brain, and the number of incident cases of schizophrenia in the population of a health authority. Although the counts observed over a fixed time period are the data to which the Poisson distribution applies, the interest is in estimating the rate that has produced them. The distribution underlies typical epidemiological methods used in incidence and prevalence studies such as Poisson regression. The distribution is characterized by the mean rate (with variance equal to the

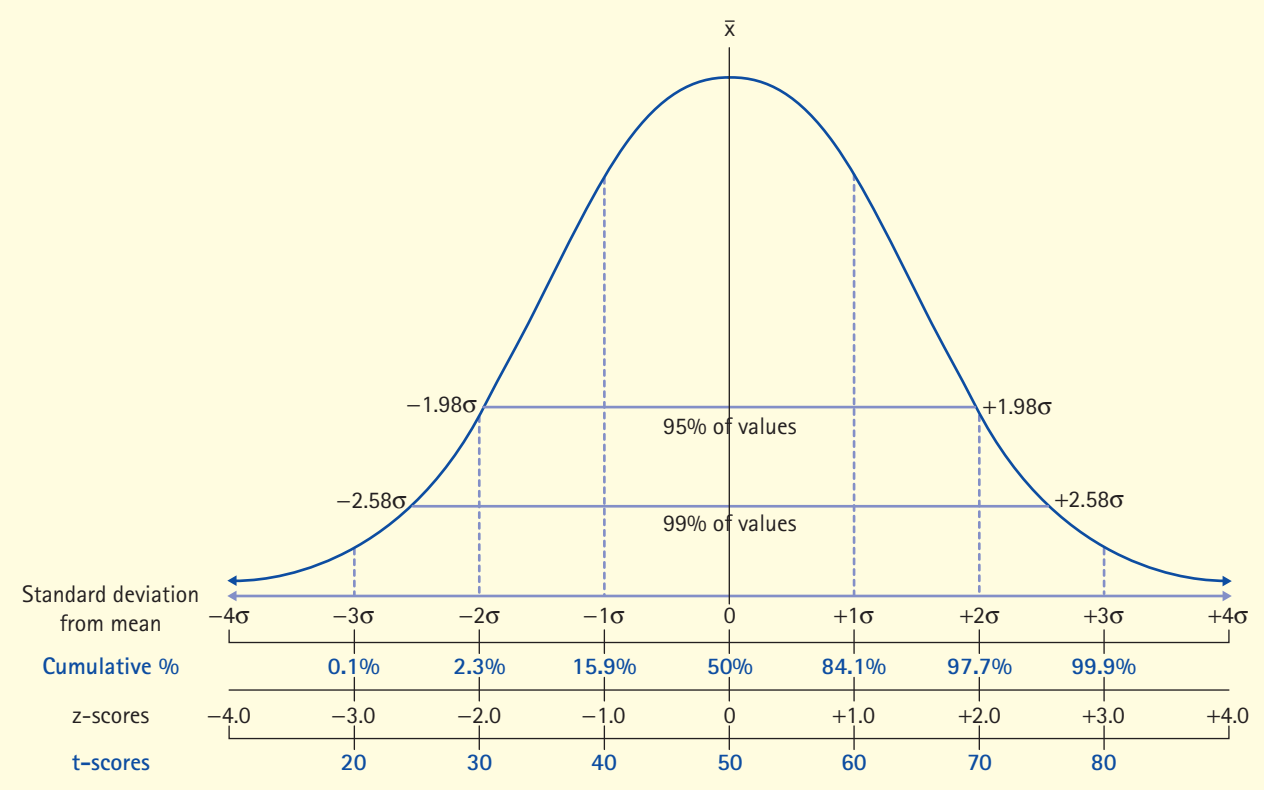

Figure 4.7 Normal or Gaussian distribution

mean), and it is approximated by the normal distribution as the rate increases (Figure 4.8).

A Bernoulli trial is a term describing a single event, for example admission of a patient to hospital within a given week. Given a fixed probability p that this event does or does not happen, the distribution of the number k of such events out of a fixed total n is described by the *binomial* distribution. The binomial distribution underlies the analysis of proportions (including logistic regression). The mean and variance of the binomial are np and $np(p-1)$. The observed proportion of events k/n estimates p. When p is small (say, < 0.05), $1-p$ is close to 1 and $p(1-p)$ is close to p. In this case, the binomial is close to a Poisson distribution with rate p. For example, if 1 person out of 10 000 dies over the course of a year, one could think of this either as a rate of 1 per 10 000 person-years or as a proportion of 1/10 000 (for a 1-year follow-up). The binomial distribution is approximated by a normal distribution in large samples unless p is very small. In both the Poisson and binomial distributions, the counts or events are assumed to be independent. In practice, this means that the probability of an event does not depend on what has happened before or, in the case of Poisson counts in a region, in the surrounding region.

Summary statistics calculated from data

For categorical data, the usual summary statistics are simply frequencies n and proportions p (expressed as a percentage) for each category. The odds $O = p/(1-p)$, where p is a proportion, is a quantity used as an intermediate quantity in analysis, for example as illustrated above for diagnostic tests or via the odds ratio in epidemiology, but it is seldom presented as a simple summary statistic. Note that the mode may be used as a (not very useful) measure of the typical value in a categorical variable. For example, if ethnic group is recorded as white (70%), black Caribbean (15%), black African (5%) and other (10%), then the modal category is white.

For rates, it is usually important to give the number of events on which the rate is based and the exposure. The term 'exposure' is slightly confusing, because it can also mean exposure to a specific risk factor; in this context, it means the total population at risk or the total follow-up time accumulated over all those studied (e.g. in a cohort study). For example, the incidence rate of schizophrenia might be expressed as 25 cases over a follow-up period of 3567 person-years (e.g. 1000 people followed for an average of 3.567 years each), giving a rate of 0.007, or 7 per 1000 person-years. Standardized mortality ratios (SMRs) provide a way of comparing rates across different populations with different characteristics (e.g. age, gender). For example, one might be interested in the rate of schizophrenia among recent immigrants where a confounding factor might be their typically younger age distribution compared with other subgroups or the population at large. A standard population is chosen (e.g. that of the UK) and is divided up into bands (e.g. by age and gender). Rates of disease (e.g.

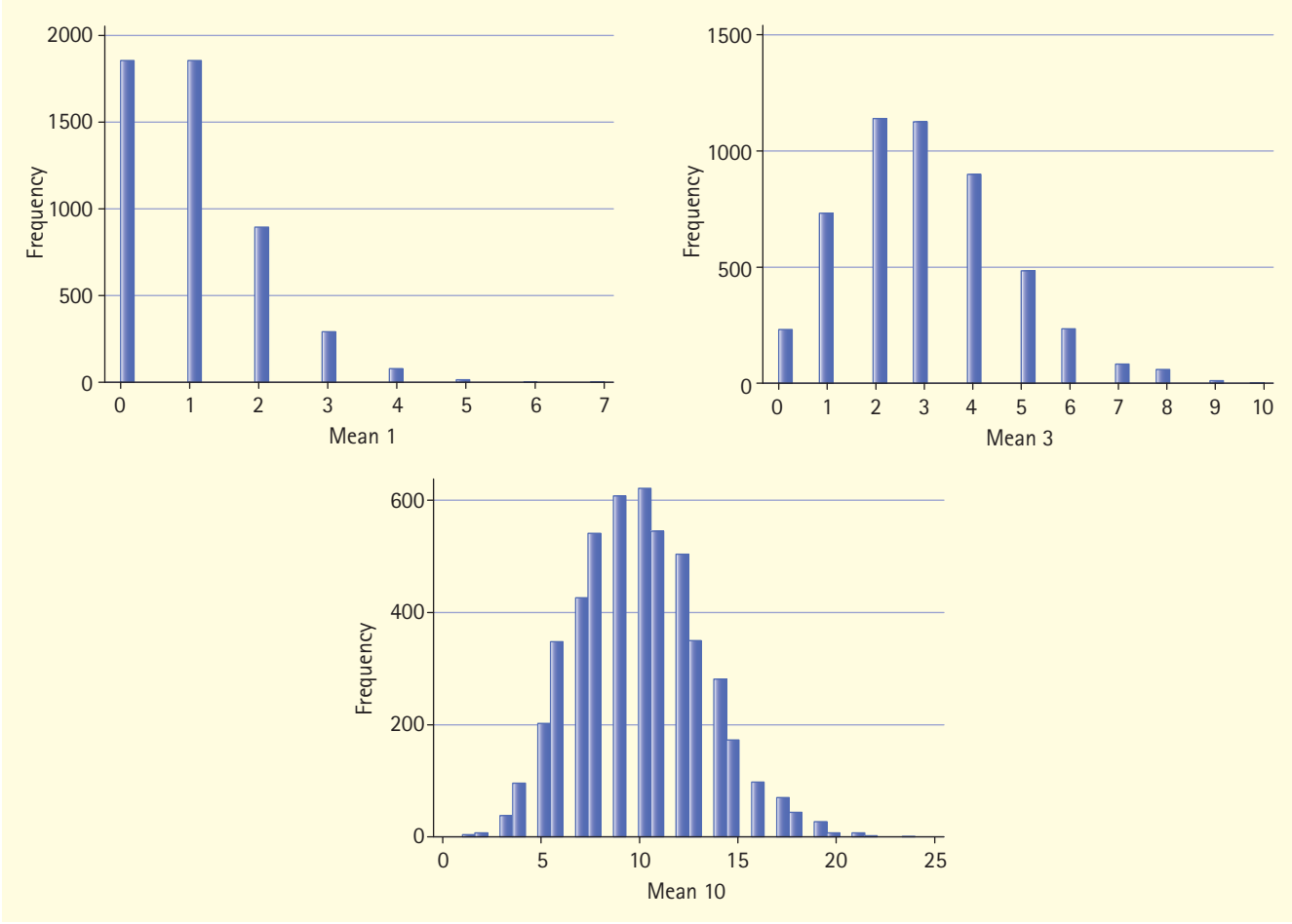

Figure 4.8 Examples of Poisson distributions (means 1, 3 and 10)

schizophrenia) in the population are found, for example from national records held by the Office for National Statistics, and are applied to the age and gender bands in the study sample. From this, the expected number of cases can be calculated for each band and accumulated to give the total expected, E. The ratio $O/E \times 100$ is the SMR, where O is the number actually observed.

For continuous data, there are a number of basic summary statistics, of which the mean, standard deviation and range, as shown in the Stata output for the change in knowledge scores for nurses in the scenario, are common:

```
Variable |   Obs      Mean    Std. Dev.     Min      Max
---------+--------------------------------------------------
  change |    30       2.8    2.483046      -2        8
```

Medians and percentiles are used to summarize skewed data (the median is equal to the mean in symmetrically distributed data but lower if the data are skewed with a tail to the right). The *mode* is the value corresponding to the peak of the distribution. Table 4.8 gives the formulae.

In papers, percentages should generally be quoted without decimal places, except in very large samples, and they are usually given in relation to the valid (non-missing) cases. The total sample size should be given with and without missing values. Means and standard deviations are usually given to a small number of decimal points (two or three). Note that confidence intervals or standard errors for means as descriptive data are usually not necessary. Rather, the important figure is the standard deviation, since this describes the variability of the data. (Confidence intervals are more useful for derived statistics from which inferences are to be made). Altman and Bland give some general guidelines on these issues.[1]

Graphical presentations

Plotting data should be the first step in any analysis. Categorical data can be displayed in bar charts and pie charts. Examples of box plots, histograms and QQ plots for two types of continuous data are illustrated in Figure 4.9. The box in a *box-and-whisker plot* (or box plot, for short) shows cases between the upper and lower quartiles. The *median* is indicated on the box plot as a bold line. It is often

Table 4.8 Basic formulae for summary statistics based on continuous data

Summary statistic	Formula (for a set of *n* data points $x_1, x_2 \ldots x_n$)	Notes
Mean	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$	Most commonly used measure of the location of data for approximately normally distributed data
Variance	$v = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$	Measure of the spread of the data around the mean
Standard deviation (SD)	$s = \sqrt{v}$	Expresses the variation in the same units as the data; the data deviate on average $\lvert s \rvert$ from the mean
95% reference range	$\bar{x} \pm 1.96 \times \text{SD}$	If data are distributed approximately normally, 95% of the individual observations lie within the reference range
Median	$\frac{(n+1)th}{2}$ value of the ordered observations	Midway value: half the data points lie below and half above the median; for an even number of cases, the average of the two middle values is used
Lower quartile (25th percentile)	$\frac{(n+1)th}{4}$ value of the ordered observations	The interquartile range is the range lying between the lower and upper quartiles; other percentiles can also be calculated
Upper quartile (75th percentile)	$\frac{3 \times (n+1)th}{4}$ value of the ordered observations	

taken as a 'typical' value, especially if the data are not distributed symmetrically. The whiskers on the plot (the thin lines) indicate the overall range of the data, after excluding very atypical values, which are known as *outliers*. The histograms show *frequency distributions*, where the range of the data are divided up into equal-sized 'bins' (usually 15–20) and the number on each bin counted and plotted. The *y*-axis scale here is frequency, but proportions can also be shown (in which case, it is important to give the total sample size). A theoretical normal probability distribution with the same mean and standard deviation is superimposed; here, the *x*-axis is continuous and the *y*-axis is a probability.

The QQ probability plots are shown on the right. They provide a way of visually estimating how close a distribution is to being normally distributed, as an alternative to formal tests such as the Kolmogorov-Smirnov test. If the data are normal, the points should lie on a straight line, as is the case for the first row of data. The second row of Figure 4.9 shows an example of a non-normal skewed distribution. (Duration of illness is often log-normally distributed.) *Skewness* refers to lack of symmetry and may be important to consider. (*Kurtosis* is another feature of non-normality, where the slope away from the peak is not so marked as for a normal distribution.) This latter feature of data is, however, rarely of any concern to data analysis. Here, the data have a tail to the right, with a few very high values, and the median is lower than the mean. Taking logs of such data sometimes produces a normal distribution (when the original distribution can be described as *log-normal*). Analysis can be carried out on the log-transformed values, and summaries such as the mean are then back-transformed.

Reference ranges and outliers

In medicine, the mean ± 1.96 standard deviations is sometimes known as the *reference range*, since it indicates where the vast majority of cases are expected to lie. It is not the same as the 95 per cent confidence interval (see Statistical inference, below). Values outside that range are not necessarily inconsistent with the distribution – they may be just at one end of a continuum. *Outliers*, on the other hand, are cases that are considered to be so extreme as to be unlikely to be part of the distribution, either because they derive from another underlying population that has been mixed in with the main population or because of a data measurement or transcription error.[24] The question of outliers is complex, but general advice would be to remove data only if there is definite evidence for a data error.

Dummy variables

In order to include the effects of categorical explanatory variables, many statistical models employ *dummy variables*

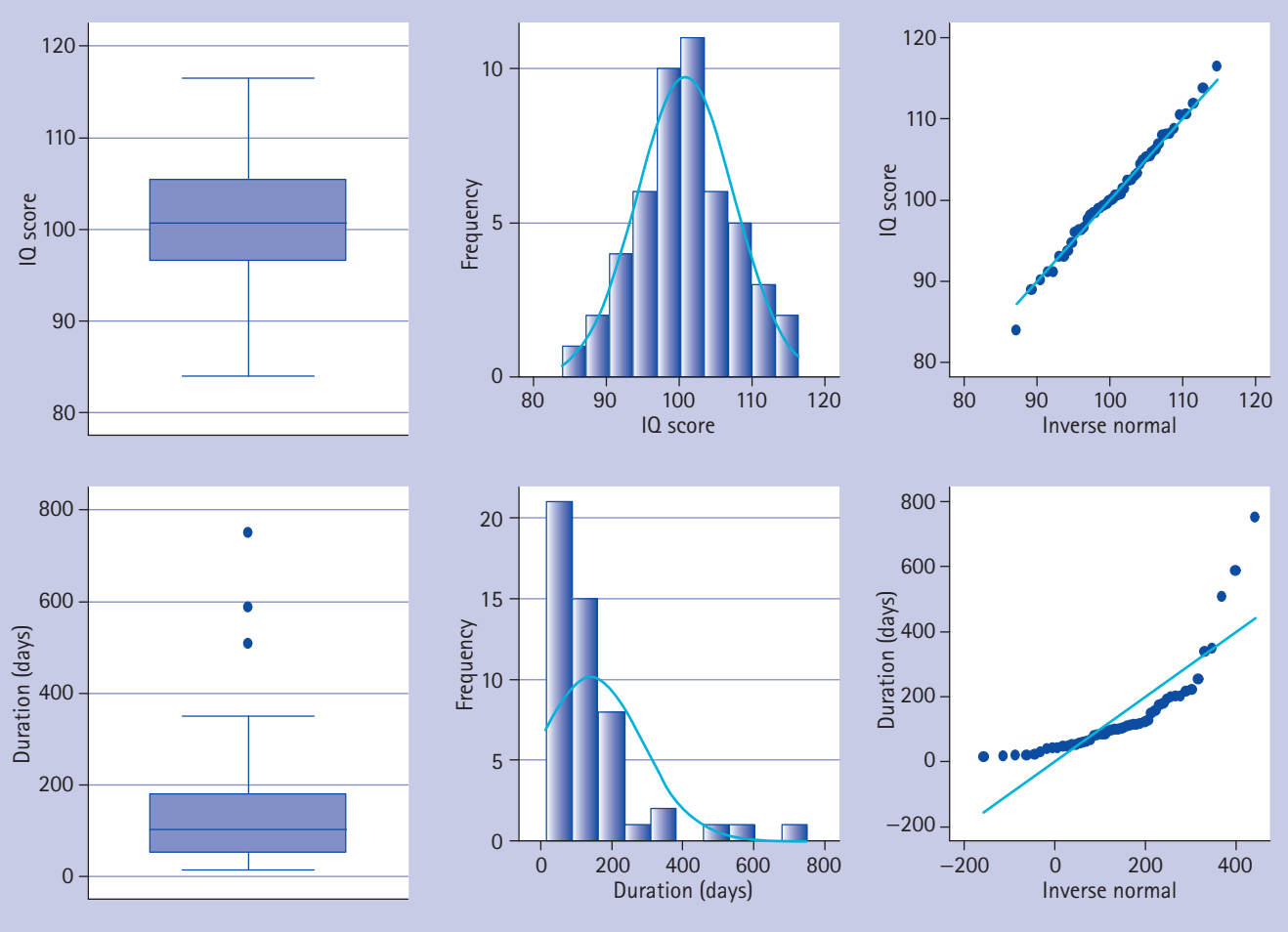

Figure 4.9 Box plots, histograms and QQ plots of a normally distributed variable (top row: intelligence quotient (IQ) of 50 patients) and a skewed, log-normally distributed variable (bottom row: days in hospital of 50 patients)

– binary indicator variables where the value 1 indicates that a subject falls into a specific category and the value 0 that it does not. Categorical variables with k categories can be represented by k dummy variables. Since the information provided by a factor with k categories is contained in k – 1 dummy variables (once you know the value of all but one, the value of the last one can be inferred), only k – 1 dummy variables are used in models. The category that has been left out then becomes the reference category.

Here is an example of dummy coding for a factor with three categories (the single variable Factor is replaced by three new variables Dum1, Dum2 and Dum3, only two of which would be included in an analysis).

Factor	Dum1	Dum2	Dum3
1	1	0	0
2	0	1	0
1	1	0	0
3	0	0	1
3	0	0	1
2	0	1	0

Effect sizes

The term ‘effect size’ often refers to the size of a mean difference or whatever quantity is the focus of interest. A measure of relative effect size can be presented to allow the scientific or clinical importance of the observed effects to be assessed on a common scale. The need for this arises because (i) a statistically significant difference in a large sample might be small and of no scientific importance; and (ii) different variables to be compared might be measured in different units. A widely used relative effect size is *Cohen’s d*, which, for the comparison of two groups, is the mean difference between two groups divided by the pooled standard deviation.[25] The *standardized response mean* (SRM) may be more appropriate as a measure of responsiveness to change.[26] It is the mean of the change scores divided by the standard deviation of the change scores. Values for Cohen’s d or SRMs between 0.2 and 0.5 would be regarded as small, between 0.5 and 0.8 as medium, and above 0.8 as large. It is important to explain how any such standardization has been performed and to be aware that using the standard

deviation computed from a particular sample may limit the generalizability of the results. Because relative effect sizes can be compared across studies, they allow the accumulation of knowledge: a special statistical technique, meta-analysis, allows the combination of the results of several studies (see Multivariate and other more complex techniques, below).

STATISTICAL INFERENCE

Point estimates, standard errors and confidence intervals

Statistical inference is concerned with drawing conclusions from data and generalizing them to populations. It typically involves calculating point estimates and estimates of uncertainty such as confidence intervals, and performing hypothesis tests (e.g. comparing two or more groups, perhaps two arms in an RCT or those at risk and those not in a cohort study).

A *point estimate* is a single estimate such as a mean. In the Stata output for the nurse example below, the mean of the change from baseline to follow-up is 2.8 (point estimate). The standard error (SE) is an estimate of uncertainty in the point estimate and is the standard deviation of the distribution of the means (here, it is $0.453 = 2.483/\sqrt{30}$, 2.483 being the standard deviation (SD) of the raw data as shown earlier).

Variable	Obs	Mean	Std. Err.	[95% Conf.	Interval]
change	30	2.8	.4533401	1.872815	3.727185

Confidence intervals (here, 1.87 to 3.73) are derived from standard errors (and they both give equivalent information), but the confidence interval is the preferred measure of the precision for a parameter estimate. In recent years, journals have increasingly recommended authors always to report confidence intervals, not least when results are inconclusive, because they indicate the largest and smallest estimates consistent with the data; these estimates are helpful in planning further research.[12,13]

What do confidence intervals show? If we repeat an experiment, trial or any other study involving a random sample many times and calculate each time a mean and its 95 per cent confidence interval, those intervals would contain the true population mean in 95 per cent of the repeats. An intuitive way of thinking about this is that it is the range within which we expect the true parameter to be with 95 per cent confidence.

The simulation illustrated in Figure 4.10 exemplifies this. We assumed that a distribution of a change score for the nurses was distributed normally with a true mean of 2 and a standard deviation of 2, and we drew 50 random numbers from this distribution, repeated 100 times. Each time, the sample mean and its 95 per cent confidence interval were calculated. Figure 4.10 shows the results. The average sample mean from the 100 simulations was 2.04, which is very close to the true mean. At no time was it estimated at the true population value of exactly 2, but 95 of the 100 95 per cent confidence intervals do include the true mean of 2. In only five simulations (filled circles) would we have made a wrong inference about the true population mean from the confidence interval.

Formulae for standard errors and confidence intervals

Tables 4.9–4.11 give formulae for standard errors and confidence intervals for a number of point estimates. It is not necessary to remember all of them but to remember rather the following key points, which relate to approximate standard errors and confidence intervals based on large samples ($n > 50$), unless otherwise stated:

- The 95 per cent confidence interval for a point estimate is given by estimate $\pm$ 1.96 $\times$ SE.
- Standard error for a mean of a normally distributed variable = $\mathrm{SD}/\sqrt{n}$.
- In small samples, the 1.96 in the formula for the confidence interval of a mean is replaced by the corresponding value of a t distribution with $n-1$ degrees of freedom.
- Standard error for a proportion $p = \sqrt{p(1-p)/n}$.
- Standard error for a count $a = \sqrt{a}$.
- Standard error of A + B = standard error of A − B $= \sqrt{SE_A^2 + SE_B^2}$.
- Confidence limits for ratios of proportions, rates and odds are calculated on the log-transformed scale and then back-transformed.
- For samples smaller than 50, 'exact' methods using special tables or software are based on the underlying Poisson or binomial distributions.
- The larger the sample size n of a study, the smaller the standard error and 95 per cent confidence interval, and thus the more precise the estimate.

Table 4.9 gives formulae for summary statistics for a single proportion p based on k successes out of n trials or for a comparison of two proportions $p_1 = a/(a+c)$ and $p_2 = b/(b+d)$ (see Table 4.3 above for the layout of data assumed for these formulae). Ratios of proportions and odds, which are commonly used in case–control studies, are natural log-transformed before standard errors and confidence limits are calculated and are then back-transformed by exponentiation. For samples smaller than 50, 'exact' methods using special tables or software are based on the underlying binomial distribution.

Table 4.10 gives the formulae for a rate based on a events with a population or follow-up time (time at risk) in terms of person-years of N, or a comparison of two rates a/N_1 and b/N_2. For samples smaller than 50, 'exact' methods using

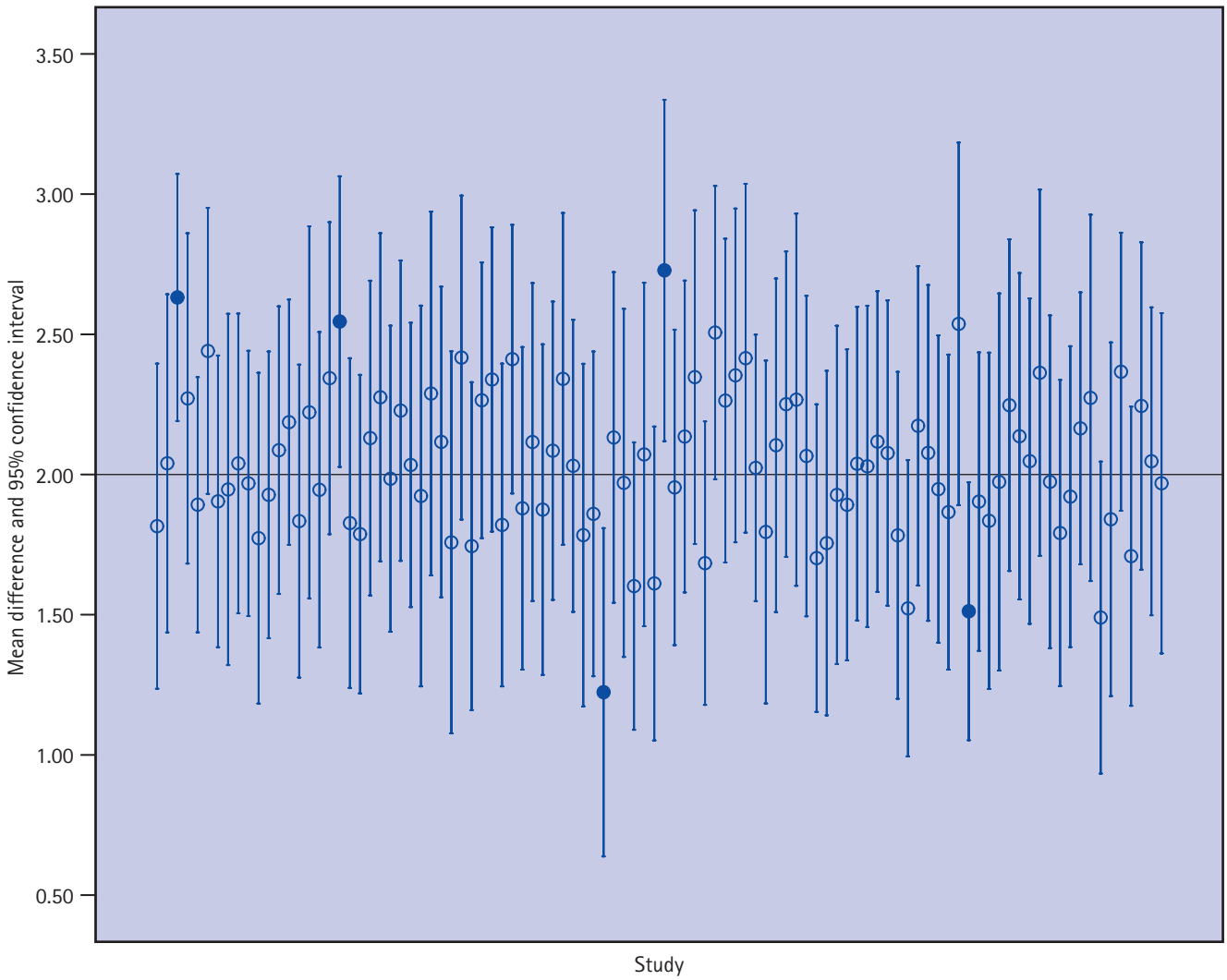

Figure 4.10 Results of 100 simulations from a distribution with mean 2 and standard deviation 2. Sample means and 95 per cent confidence intervals computed from 50 cases each time

special tables or software are based on the underlying Poisson distribution for counts.

The standard errors and confidence intervals for continuous data that can be assumed to be approximately normally distributed are shown in Table 4.11. For large samples (> 50), the 95 per cent confidence interval is given by estimate $\pm 1.96 \times$ standard error; for small samples, the appropriate point of the T distribution should be used instead of 1.96.

Note that in the above formula for the variance, the denominator is $n - 1$ and is called the *degrees of freedom*. It can be shown that the estimate of the underlying population variance is unbiased if $n - 1$ is used as a denominator rather then n.

Hypothesis tests

In many studies there is a prediction of a difference between groups. This prediction is variously known as the experimental, research or *alternative hypothesis* (H1). The converse hypothesis is the *null hypothesis* (H0). We test H0 to see whether we can reject it beyond reasonable doubt in favour of H1. The procedure of classical hypothesis testing is to define a research and a null hypothesis, to collect sufficient data and to calculate a test statistic z^*; this is compared against a probability distribution, which is known if the null hypothesis is true. We reject H0 if the observed test statistic is very unlikely to occur if H0 is true, according to the probability that $Z > z^*$, where Z is a theoretical random variable with the known distribution. A null hypothesis is rejected if this probability – the P value – is below a *critical value*, α; thus, P is not the probability that H0 is true but the probability of an extreme value if it is true.

Typical values for α are 0.1, 0.05 and 0.01. They may be adjusted if multiple tests are to be performed using, for example, the Bonferroni correction: this would divide α by as many tests as are to be applied simultaneously (e.g. see Shaffer[27] for a discussion of multiple testing). In the vast majority of cases, one assumes that the null hypothesis

Table 4.9 Standard errors and confidence intervals for summary statistics based on binary data (proportions)

Standard error (SE) of proportion *p*	$\sqrt{\frac{p(1-p)}{n}}$	
Standard error for difference between two proportions	$\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}$	When p_1 and p_2 are estimates of the same proportion (e.g. when two groups do not differ), the pooled estimate is used for both p_1 and p_2: $p=\frac{n_1p_1+n_2p_2}{n_1+n_2}$
Risk ratio (RR)	$\frac{p_1/n_1}{p_2/n_2}$	An alternative to risk difference
Standard error for ln(RR)	$\sqrt{(1/a+1/b+1/n_1+1/n_2)}$	Compute the CI for ln(RR), then exponentiate (anti-log) the limits for RR
Odds ratio (OR)	$\frac{p_1/(1-p_1)}{p_2/(1-p_2)}$ or equivalently $(a\times c)/(b\times d)$	An alternative to risk difference or ratio; produced by logistic regression
Standard error for ln(OR)	$\sqrt{(1/a+1/b+1/c+1/d)}$	Compute the CI for ln(OR), then exponentiate (anti-log) the limits for OR

CI, confidence interval, *a* and *b*, frequencies (see Table 4.3).

Table 4.10 Standard errors and confidence intervals for summary statistics based on count data (rates)

Standard error (SE) of rate	$\sqrt{\frac{a}{N_1}}$	Uses the variance = mean relationship for Poisson variables
Rate ratio	$\frac{a/N_1}{b/N_2}$	
Standard error for ln(ratio of counts *a* and *b*) to find confidence interval for rate ratio	$\sqrt{(1/a+1/b)}$	Compute the CI for ln(count ratio), then exponentiate (anti-log) the limits and multiply by N_2/N_1 to convert to rates

CI, confidence interval, *a* and *b*, counts N_1 and N_2, follow up times.

would be rejected with an extreme finding in either direction (i.e. large or very small), so a *double-sided test* is applied. Occasionally one knows in advance that deviations are impossible in one direction, so a *single-sided test* is applied.

In a single-sided test, α corresponds to the critical value z^*, such that:

The probability that $Z > z^* = \alpha$

For example, if the desired significance level is 0.05, testing a mean value assuming a normal distribution for Z, the corresponding value for z must be greater than or equal to 1.645 (or less than or equal to −1.645) for a single-sided test. For a double-sided test (where the extreme values can be at either end of the distribution), we are interested in:

The probability that $Z > z^* = \alpha/2$

Table 4.11 Standard errors and confidence intervals for summary statistics based on continuous data (assumed normally distributed)

Standard error for mean $\bar{x}$ (SE)	$\frac{s}{\sqrt{n}}$	s is SD; n is sample size
95% confidence interval for mean $\bar{x}$	$\bar{x} \pm 1.96 \times SE$ or $\bar{x} \pm t^* SE$	*If the total sample size is small (< 50), t^* is the 2.5% point of the t distribution with $n_1 - 1$ degrees of freedom
Standard error difference in two means $\bar{d}$	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1^2 + n_2^2 - 2}}$ A pooled estimate of s can be used for s_1 and s_2 if they are not significantly different
95% confidence interval for difference between two means $\bar{d}$	$\bar{d} \pm 1.96 \times SE_d$ or $\bar{d} \pm t^* \times SE_d$	*If the total sample size is small (< 50), t^* is the 2.5% point of the t distribution with $n_1 + n_2 - 2$ degrees of freedom

so that probability is divided up between the two extremes of the distributions and the critical value z^* corresponds to the $\alpha/2$ significance level. To achieve a significance level of 0.05 for a double-sided test, the absolute value of the test statistic ($|z|$) must be greater than or equal to the critical value 1.96 (which corresponds to the level 0.025 for a single-sided test).

A 95 per cent confidence interval can also be used as a criterion to reject a null hypothesis: if it does not encompass the null hypothesis value, then the effect is statistically significant at $P = 0.05$. The null hypothesis value would be 0 (for a comparison of means in two groups) or 1 (for a comparison of proportions in two groups using ORs or RRs). A test based on P values and a test based on confidence intervals are closely related, and they will almost always result to the same conclusion about rejection of a null hypothesis.

Power and sample size calculations

When performing significance tests using $\alpha = 0.05$, over the long run the rejection of H0 will be mistaken in 5 per cent of occasions; this is a *type I error*. The probability of accepting the null hypothesis even though the alternative hypothesis is in reality true is denoted by β and $1 - \beta$ is known as the *power* of a test. The error is known as a *type II error*. The latter is more difficult to assess, but it is possible to control for it by doing a power analysis during the planning stage of a study. Here is a summary:

- α-*error:* probability of making a type I error (rejecting H0 although H0 is true).
- β-*error:* probability of making a type II error (accepting H0 although H1 is true).

A sample size calculation requires the following information:

- The power for the particular trial: $1 - \beta$, where β is the probability of a type II error – typically 80 per cent or 90 per cent
- The smallest difference (or 'effect') that is considered to be of clinical importance
- The significance level α, or probability of a type I error, to be used – typically 0.05
- The expected level of loss to follow-up.

For continuous data, one needs an estimate of the standard deviation within groups; if this is unknown, then Cohen's standard effect size (see above), which by definition has a standard deviation of 1, can be used as a generic solution. Sample sizes can be calculated using specialized computer software or using formulas and tables from standard textbooks (e.g. Cohen[25]). A rule of thumb for the sample size for a power of 80 per cent and a two-sided $\alpha = 0.05$ for independent t-tests or χ^2 tests is:

> Sample size per group = $16/d$, where d = Cohen's d (the standardized difference)

Assumptions made by tests

Many statistical tests for continuous data depend on assumptions, for example that the data are distributed normally or that the variances of different groups are the same (*variance homogeneity*). Tests that depend on specific distributions are described as *parametric*. Before commencing with the main analysis, one may wish to investigate these assumptions. Some formal tests are available (e.g. Bartlett test for normality, Levene test for variance homogeneity), but they are often not helpful because datasets with small sample sizes will not reject the null hypothesis and datasets with large sample sizes may reject the null hypothesis even if the violations are minor and have little impact on the conclusions. Equality of variances can be evaluated by a rule of thumb: the larger standard deviation should be less than twice the smaller standard deviation.[28] Normality is best assessed visually by plotting box plots, histograms or QQ plots of the observed data for each group.

A further assumption that may be made is of independence of observations. This is often invalid for psychiatric

data and then special methods are required, either robust methods or methods specially designed for non-independent or clustered data (see Multivariate and other more complex techniques, below). *Robust methods* include those that use standard models, such as regression, but with adapted variances for the parameter estimates (so-called *sandwich estimates*), with the aim of more realistic *P* values even when the distribution of the data is not as specified in the model. Bootstrapping is another robust method that works similarly to simulation (see section on health economics, where bootstrapping is commonly used).

Transformation

If the assumptions of a normal distribution and/or equal variances are not satisfied, then it may be advisable to transform the data. In psychiatric studies, data are often positively skewed, especially with essentially positive values such as service-use or cost data. Log-transformation is helpful both for producing normal distributions and also for unequal variances if the larger variance is associated with higher values. Note, however, that the back-transformed mean of a lognormal distribution is the geometric mean and is closer to the median of the original values rather than to the arithmetic mean. The arithmetic mean of the original untransformed data is estimated by $\exp(m + s^2/2)$, where m and s are the mean and standard deviation of the (natural) log-transformed values, respectively. Other transformations are the square transformation (x^2), which is used if the data are negatively skewed and the square root ($\sqrt{x}$), which is recommended for count data (although note that it is usually preferable to use methods based on the Poisson distribution rather than to transform and assume normality).

Non-parametric tests

An alternative to transformation is to use non-parametric tests, which do not assume particular distributions and are usually based on ranks. They do make other assumptions, however, depending on the test. A disadvantage of non-parametric tests is the difficulty of deciding on the appropriate summary statistic to illustrate group differences (instead of the mean and standard deviation). Another disadvantage is that it is difficult to extend analyses to take account of confounders (relatively easy with normally distributed data using regression). Siegel and Castellan[29] is a good basic text on non-parametric tests.

What do significant and non-significant results mean?

It is important to remember that statistical significance does not tell us anything about the importance of the effect. It reflects the amount of evidence that the effect exists. Small and clinically unimportant effects can become statistically significant if the sample size is large enough. Two quantities mentioned above are helpful in balancing statistical and clinical significance: the 95 per cent confidence interval for the point estimate, which quantifies the possible range of the effect, in original units, that is consistent with the data, and the standardized effect size, which places the point estimate on a scale measured in units of standard deviations.

As Altman and Bland discuss, 'absence of evidence is not "evidence of absence"'.[30] If the *P* value is above 0.05, this tells us that the observed effect is not large enough to exclude chance as an explanation. Thus, if there is no evidence to reject the null hypothesis, this does not mean that we have proved it to be true. Classical or *frequentist* hypothesis testing is thus fundamentally asymmetrical in that more weight is given to the null hypothesis, since it is accepted unless there is overwhelming evidence to reject it.

The Bayesian approach

The frequentist approach may seem not to address the aim of the study, which is to assess the probability that a hypothesis is true. Furthermore, this approach uses only information from the current study and does not integrate other available information, such as results of previous studies or opinions of psychiatrists. The *Bayesian* approach adopts a different viewpoint, more concerned with weighing the evidence both for and against a research hypothesis. It is based on the beliefs (about parameters and the hypothesis) as well as the data (see Bolstad[31] for an accessible introduction). In Bayesian statistics, population parameters are considered to be random variables and thus have probability distributions. In this context it is possible to express the belief or the probability that a research hypothesis is true. The belief in the hypothesis before the study starts is called the *prior probability*, which is converted to a *posterior probability* by the analysis of the data. This update is based on conditional probability through *Bayes' theorem*. A major criticism of Bayesian statistics is the subjectivity of describing prior belief and, in the past, the difficulty of the mathematics and computation involved. However, Bayesian methods have become increasingly popular, especially in decision-making analysis.[32]

ILLUSTRATIONS OF HYPOTHESIS TESTS

Summary of tests

Table 4.12 summarizes the most important parametric and non-parametric tests and when they are used. The most commonly used tests will be illustrated later in this section.

The statistical distributions used in the tests in Table 4.12 are the Student's *t*, χ^2 and F distributions. Their distributions depend on the *degrees of freedom*, which is generally

Table 4.12 Commonly used hypothesis tests

Comparison being made	Scale of measurement	Test
One group (one sample mean against population mean or distribution *v.* theoretical distribution)	Nominal	χ^2 goodness-of-fit test
	Ordinal	Mann–Whitney U test
	Interval/ratio	One sample *t*-test
Two groups (independent samples)	Nominal	χ^2 test, Fisher's exact test
	Ordinal	Mann–Whitney U test
	Interval/ratio	Two-samples *t*-test
Two groups (related samples)	Nominal	McNemar test
	Ordinal	Wilcoxon signed rank test
	Interval/ratio	Paired sample *t*-test
More than two groups (independent samples)	Nominal	χ^2 test for homogeneity
	Ordinal	Kruskal–Wallis test
	Interval/ratio	Analysis of variance (ANOVA)
	Survival	Log-rank test

the sample size minus the number of parameters that have been estimated (e.g. for the *t* distribution there are two parameters, the mean and standard deviation). Table 4.13 shows how these distributions are usually represented (the *d*s here are degrees of freedom, and some relationships that are worth remembering). The specific tests will be discussed later.

The independent samples *t*-test

Let us illustrate hypothesis testing with an example of the knowledge of nurses before and after a training programme. We are interested in whether nurses with training improved their knowledge more than nurses on the waiting list, on the basis of the change in knowledge test scores. What is the likely range of the true mean difference in change scores, and is there evidence that this difference is significantly different from the null hypothesis value of 0? The test statistic is the mean difference between the two groups divided by its standard error – the *t* statistic. The larger the difference between the two groups and the smaller the standard error, the larger the (absolute) value of *t*. The size of *t* is judged in relation to the theoretical *t* distribution; if it is above a critical value corresponding to $\alpha = 0.05$, it is deemed to be significant.

The output below presents the *P* value, which is 0.0018. The chance of obtaining a *t*-value greater than |3.45| if the null hypothesis is true is thus less than 0.05 and we can therefore reject the null hypothesis. Alternatively, we can say that the observed mean difference of improvement

Table 4.13 The main distributions used in statistical tests

Distribution	Used for	Equal to
χ^2_d	Comparing proportions, as a fit index and for some non-parametric test statistics	Distribution of sum of squares of $N(0,1)$ variables
t_d	Comparing means (one mean against a known value or two groups compared with each other)	$N(0,1)$ distribution, when n is large
$F_{d1,d2}$	Testing whether the ratio of variances is 1 and for comparing means (when there are more than two, in an analysis of variance)	t_{d2} distribution, when $d_1 = 1$

```
ttest   change, by(group)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
 Waiting |      15    1.466667    .4457916    1.726543    .5105388    2.422795
Training |      15    4.133333    .6314508    2.445599    2.779006    5.487661
---------+--------------------------------------------------------------------
combined |      30         2.8    .4533401    2.483046    1.872815    3.727185
---------+--------------------------------------------------------------------
    diff |           -2.666667    .7729556               -4.249994   -1.083339
------------------------------------------------------------------------------
    diff = mean(Waiting) - mean(Training)                         t =  -3.4500
Ho: diff = 0                                     degrees of freedom =       28

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0009         Pr(|T| > |t|) = 0.0018          Pr(T > t) = 0.9991
```

between waiting list and training group nurses of −2.67 is too large to be explained by chance. The 95 per cent confidence interval of the difference in change score is −4.25 to −1.08 and does not include 0. It also informs us that the test can be rejected at 0.05. Additionally, it tells us that the possible mean difference ranges between −4.25 and −1.08 (see Stata output, above).

The *t*-test assumes that in the population the variable is distributed normally in each group, the variances are the same (variance homogeneity) and the observations are independent of each other. The *t*-test is fairly robust against some violations of the assumptions of normal distribution and variance homogeneity but not against violations of the independence of observations. As *n* increases, the *t* statistic approaches the *z* statistic, which has a $N(0,1)$ distribution. For large samples, therefore, any estimate of a mean difference from an approximately normal distribution whose absolute value is more than twice its standard error is significantly different from 0 at $P = 0.05$. This rule of thumb comes from the (rounded) double-sided 0.05 point of the normal distribution i.e. 1.96.

The Mann–Whitney U (Wilcoxon rank sum) test

A non-parametric test that can be used for comparing two independent groups when the data are not normally distributed is the Mann–Whitney U test (MWU; also known as the Wilcoxon rank sum test). The MWU test is based on the ranks of the data and does not have any distributional assumption. However, note that it tests equality of medians only if the distributions have the same shape. Figure 4.11 shows two types of non-normal data, illustrating how

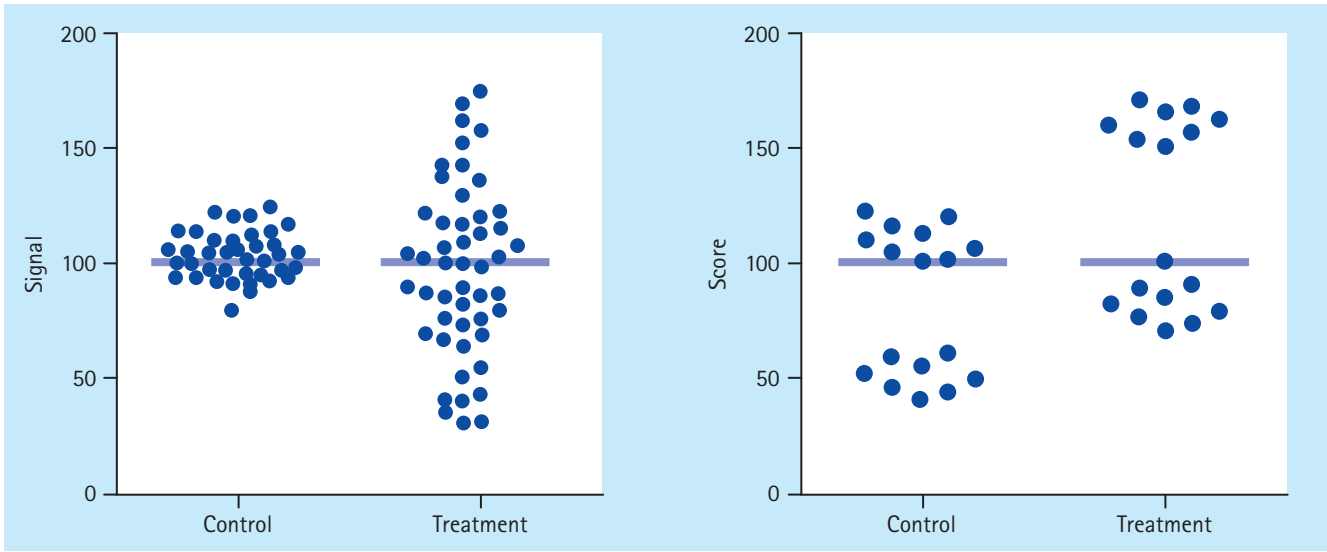

Figure 4.11 Examples showing medians that are the same but where the distributions differ and vice versa. The Mann–Whitney test would be significant in the first example but not the second

```
ranksum change, by(group)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

       group |      obs    rank sum    expected
-------------+---------------------------------
     Waiting |       15         157       232.5
    Training |       15         308       232.5
-------------+---------------------------------
    combined |       30         465         465

unadjusted variance      581.25
adjustment for ties      -18.36
                         ----------
adjusted variance        562.89

Ho: diff_k~e(group==Waiting) = diff_k~e(group==Training)
              z = -3.182
    Prob > |z| =   0.0015
```

important it is to plot the data to decide whether the median interpretation is reasonable.

Using the MWU test for our example, we get the following output (see Stata output, above).

The test statistic is $z = -3.182$, $n_1 = 15$, $n_2 = 15$, $P = 0.0015$ (based on an approximation), so at the 0.05 level we can reject the null hypothesis that the two groups have the same distribution in the population. If the sample size is small (< 10), exact P values should be requested from the statistical software or the test statistics should be referred to statistical tables.[29]

Comparing more than two groups: analysis of variance

Analysis of variance (ANOVA) is used to test differences among two or more means. It is closely related to linear regression and can be analysed using regression techniques; the t-test is a special case. The F-test in an ANOVA compares the variability between means to that within means. Once an overall difference is established, post-hoc comparisons between groups can be applied using either specific predefined contrasts or all pair-wise comparisons with a suitable adjustment for multiple testing (e.g. Bonferroni adjustment or the Tukey test, which is less conservative). In the example below, the F-test indicates strong evidence for differences among the groups ($P = 0.0009$). The regression-style output below the F-test shows differences between dummy-coded variables (see Statistical inference, above) with the reference category, the highest-coded group in this case 3, denoted 'dropped'. Here, group 1 is not significantly different from group 3 but group 2 is (see Stata output, below).

Here is a contrast, that between groups 1 and 2, which

```
  anova gaf  group3,regress

      Source |       SS       df       MS              Number of obs =      60
-------------+------------------------------           F(  2,    57) =    7.94
       Model |  1920.75756     2  960.378782            Prob > F      =  0.0009
    Residual |  6895.82577    57  120.979399            R-squared     =  0.2179
-------------+------------------------------           Adj R-squared =  0.1904
       Total |  8816.58333    59  149.433616            Root MSE      =  10.999

------------------------------------------------------------------------------
         gaf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
_cons |  61.86667   2.839946    21.78   0.000     56.17977    67.55356
group3 |
           1 |   4.061905   3.519373     1.15   0.253    -2.985518    11.10933
           2 |   14.66275   3.896374     3.76   0.000     6.860391     22.4651
           3 |  (dropped)
------------------------------------------------------------------------------
```

has not been produced by the regression output above: it has to be asked for specifically:

```
test _coef[group3[1]]=_coef[group3[2]]

( 1)  group3[1] - group3[2] = 0

      F(  1,     57) =    9.83
           Prob > F =    0.0027
```

In a *factorial design*, several factors are varied in the same experiment. For example, drugs A and B may be given at two levels, to give four combinations. This design allows two questions to be answered at the same time, and in particular it can test whether the combination of A and B is more effective than the additive effects of either (a phenomenon known as synergy). This design is easily analysed using the ANOVA framework. A synergistic effect would result in a significant interaction F-test between drug type and level. Figure 4.12 shows another example of an interaction between two factors, gender and time (pre- and post-drug treatment): the blood level of a certain hormone was measured in males and females before and after a drug treatment. The effect of the drug on hormone level depends on gender: only in females was an increase of the mean blood levels of a hormone after the drug treatment observed.

Where a continuous variable is controlled for (e.g. severity of illness), ANOVA is known as *analysis of covariance* (ANCOVA); and is usually analysed by regression. The Kruskal–Wallis test is a non-parametric version of ANOVA that compares medians in two or more groups.

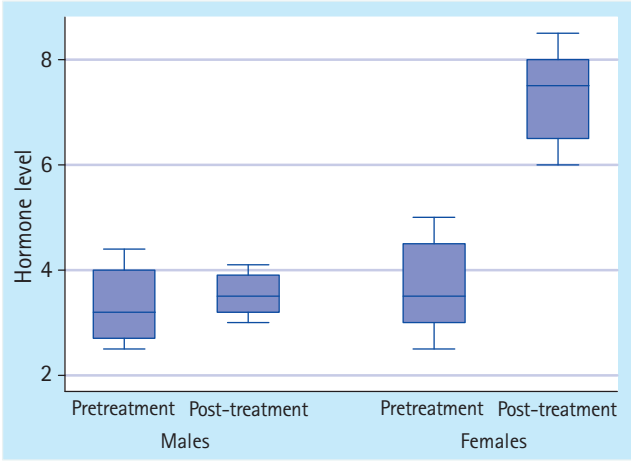

Figure 4.12 Example of an interaction between two categorical variables. The plot shows the blood levels of hormone before and after a drug treatment in males and females

Comparing two related samples: paired *t*-test and Wilcoxon signed rank test

Often we want to compare changes within a group, such as the pre- and post-training knowledge for those nurses who were trained. In this case, our two measurements are not independent. A nurse who scored above average in the knowledge test before the training started is more likely to score above average after the training programme than a nurse who scored poorly at the beginning. We need to consider the dependency of these repeated measurements, otherwise we will lose the advantage of the within-individual comparisons. This is done by the dependent samples *t*-test or its non-parametric equivalent, the Wilcoxon signed rank test. The null hypothesis here is that the mean change in the group with training is zero (see Stata output, below).

The mean difference between post- and pre-training knowledge is 4.13, with a 95 per cent confidence interval of 2.779 to 5.488. Because 0 is not included in the 95 per cent confidence interval, we can conclude that there is a significant increase in knowledge after the training. This conclusion is confirmed by the test statistic $t = 6.5458$.

A similar result is obtained if a non-parametric test is used. A Wilcoxon signed rank test shows that the pre- and

```
ttest   post_knowledge= preknowledge if group==1

Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
post_k~e |      15        17.4     .623355    2.414243    16.06304    18.73696
prekno~e |      15    13.26667    .6723567    2.604026    11.82461    14.70873
---------+--------------------------------------------------------------------
    diff |      15    4.133333    .6314508    2.445599    2.779006    5.487661
------------------------------------------------------------------------------
     mean(diff) = mean(post_knowledge - preknowledge)              t =   6.5458
 Ho: mean(diff) = 0                                     degrees of freedom =       14

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

post-training scores are significantly different ($z = 3.358$, $N = 15$, $P < 0.001$). Here, N is the sample size minus the number of tied observations within a subject. Exact P values should be calculated from tables or software if the sample size is smaller than 10.

```
Wilcoxon signed-rank test

      sign |    obs   sum ranks  expected
-----------+---------------------------
  positive |     14         119        60
  negative |      1           1        60
      zero |      0           0         0
-----------+---------------------------
       all |     15         120       120

unadjusted variance      310.00
adjustment for ties       -1.38
adjustment for zeros       0.00
                        ---------
adjusted variance        308.63
Ho: post_knowledge = preknowledge
             z =   3.358
    Prob > |z| =   0.0008
```

The above paired tests are also used if the individuals of two groups are different but linked in some other way. In twin studies, twins with a psychiatric disorder are often compared with their respective unaffected sibling. Parent–child studies and comparison between the two halves of the brain are other examples.

Note that these paired tests look only at changes within a group; if a comparison between groups is the issue, one would use an independent *t*-test on the change scores as above: two separate within-group P values are not appropriate for comparing groups (e.g. a contrast between a significant P value in one group but not in the other). This is because, as already noted, the non-significant P value indicates not absence of an effect but lack of evidence for an effect.

A single group: one-sample *t*-test

The one-sample *t*-test assesses whether the mean value differs from a theoretical value, for example does the mean knowledge score of the nurses differ from 10? The comparison value is often called a *norm* (a value obtained from very large sample, perhaps over many years, and that becomes a standard; intelligent quotient (IQ) of 100 is one such norm). Again, the confidence interval of the estimated mean of 13.5 ranges between 12.58 and 14.49 and does not include 10. Therefore, we can conclude that the nurses score significantly better than 10. A formal one-sample *t*-test below shows that we can reject the null hypothesis that the knowledge score of the nurses before the treatment started is 10 at the 0.05 level and again conclude that nurses score significantly better than 10 (see Stata output, below).

If the data are seriously skewed, we can use the one-sample Wilcoxon test as a non-parametric alternative.

Chi-squared and Fisher's exact test, and McNemar's test for paired proportions

The chi-squared (χ^2) test can be used as a *goodness-of-fit* test to compare a theoretical and an observed distribution. More widely used is the χ^2 test for proportions in a two-way contingency table. It is based on a comparison of observed and expected frequencies in the cells of the table. The degrees of freedom are $(k_1 - 1) \times (k_2 - 1)$, where k_1 and k_2 are the number of categories in the rows and columns, respectively. An amendment to the basic χ^2 statistic, aimed at improving the fit to distributional assumption, is the *continuity correction* or *Yates' correction*. If any of the expected frequencies is less than 5, then the distributional approximation to the χ^2 test statistic is inaccurate and *Fisher's exact test* would be used. This test does not work with a test statistic as such but enumerates all possible tables with the same marginal totals to find the proportion as extreme as that observed and so produces only a P value. The example below illustrates a test of whether diagnostic groups differ according in the proportions above and below a functioning threshold as measured by the GAF.

```
. ttest pre_knowledge=10

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
pre_kn~e |      30    13.53333    .4666667    2.556039    12.57889    14.48777
------------------------------------------------------------------------------
    mean = mean(pre_knowledge)                                    t =   7.5714
Ho: mean = 10                                 degrees of freedom =       29

    Ha: mean < 10                 Ha: mean != 10                 Ha: mean > 10
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

```
numerating sample-space combinations:
stage 3:  enumerations = 1
stage 2:  enumerations = 3
stage 1:  enumerations = 0

        |      primary diagnosis
        |
 gafbin | schizophr  bipolar  other | Total
        |                      psy |
--------+---------------------------+------
      0 |        24        13     10 |    47
        |      24.3      14.1    8.6 |  47.0
--------+---------------------------+------
      1 |         7         5      1 |    13
        |       6.7       3.9    2.4 |  13.0
--------+---------------------------+------
  Total |        31        18     11 |    60
        |      31.0      18.0   11.0 |  60.0

     Pearson chi2(2) = 1.4363  Pr = 0.488
      Fisher's exact =                0.587
```

In the above table, the expected values E are given underneath the observed values O. These are the average values expected in tables with no association between rows and columns and hence are proportional to the row and column proportions. The E for the schizophrenia/gafbin = 0 cell is $47 \times 31/60 = 24.3$, very close to the observed value O of 24.

In anything more complicated than a 2×2 table, the aim would be to identify the proportions that were particularly high or low, once a significant overall difference had been established. One essentially exploratory approach appropriate for large tables is to scan the *Pearson residuals* in the cells ($|O - E|/\sqrt{E}$) (any of the latter above 2–3 being worth investigating). Correspondence analysis may also help (see Multivariate and other more complex techniques, below). In smaller tables, one can test for specific (preferably prespecified) comparisons using subsets of the cells, again using a χ^2 test but bearing in mind that, with k degrees of freedom, only k possible independent tests are possible (see Everitt[33]).

The *McNemar test* is used for paired proportions. For example, one might compare the proportion of nurses who passed an exam at baseline with the number passing at follow-up after training. If there were no effect of training, the proportion changing in either direction would be expected to be 50 per cent and so the test statistic compares this with the proportion who do change in one direction, for example from fail to pass. Those who pass or fail on both occasions do not enter into the calculation.

Kaplan–Meier analysis and log-rank test

Researchers are often interested in the time until participants in a study experience a specific event such as death or recovery or relapse (e.g. time until a relapse of a psychosis). The participants are followed from a starting point (which can be different for each participant) until the time when the event of interest occurs. If the participant withdraws or the study ends before the event occurs, the data are described as *censored*: we do know that the time to event (*survival time*) is longer than the measured follow-up, but we do not know the actual time. Simple statistics such as the standard χ^2 test or the mean time to the event are oversimplistic for censored data, which are preferably analysed with *survival analysis* methods.[34]

A popular non-parametric survival method is Kaplan–Meier analysis, which allows the estimation of a survival function (proportions of people at successive time intervals who survive without the event occurring). It is drawn as a plot with time on the x-axis and survival probability on the y-axis drawn as a step function, since only times when events actually happen are plotted. The time at which 50 per cent of people survive is the median survival time, and this is often used as a summary statistic, along with its 95 per cent confidence interval. Figure 4.13 shows the survival curve for patients with severe Alzheimer's disease in two different treatment groups. The difference in survival probability increases until about 50 months. Then the curves start to merge and after 12 months both curves approach 0.

Formal tests such as the *log-rank test* are used to compare the survival distributions for two or more groups. This is a χ_1^2 test that compares the observed numbers of events at each time point with the number expected if the survival curves were the same for the two groups. It does this by ordering the survival times of all participants and hence dividing up the follow-up time into intervals in which events occur. In each time interval, the number of events is recorded and the number of participants who remain at risk is reduced accordingly. The numbers observed and expected under the null hypothesis of no difference between the groups are accumulated over the whole time period. See Collet[35] for further details.

The distribution of the times to death for the two groups is significantly different according to the log-rank test. The median time to death of patients in the new treatment group is 40 months, while it is only 27 months in the old treatment group (see Stata output overleaf).

Other survival analysis methods such as Cox regression can allow for the effects of covariates (including covariates that change over time) on the survival time.

CORRELATION AND REGRESSION

So far, we have analysed measurements from different groups. Often, however, we want to assess the relationship between two variables in a single group. For example, using the nurse dataset, one might wish to assess the relationship between the knowledge score before and after the training program. Does a high score after training depend on a high score before training? Such an analysis should start with a scatter plot, in which one variable is plotted against a

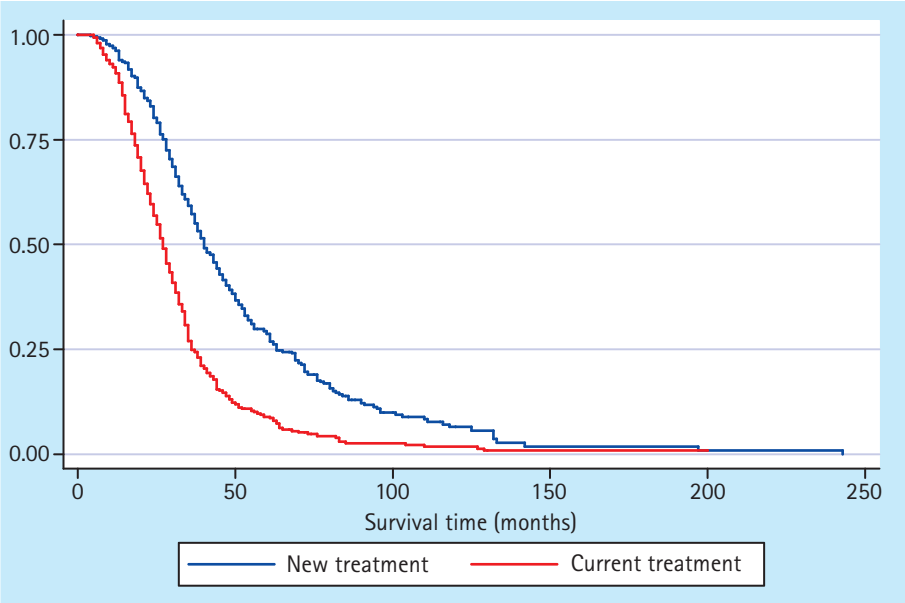

Figure 4.13 Example of Kaplan–Meier curve

```
        failure _d:  event == 1
  analysis time _t:  time

         |               incidence       no. of    |------ Survival time -----|
treatm~t | time at risk     rate        subjects        25%       50%       75%
---------+--------------------------------------------------------------------
new trea |        19695   .0181264           450         28        40        63
 current |        13437   .0305872           450         18        27        36
---------+--------------------------------------------------------------------
   total |        33132     .02318           900         21        32        50


Log-rank test for equality of survivor functions

                         |   Events         Events
treatment                |  observed       expected
-------------------------+-------------------------
new treatment method     |       357         486.23
current treatment method |       411         281.77
-------------------------+-------------------------
Total                    |       768         768.00

                           chi2(1) =     100.50
                           Pr>chi2 =     0.0000
```

horizontal axis (*x*-axis) and another one against a vertical axis (*y*-axis). If one variable is considered to be dependent on the other (e.g. if one follows the other in time), then that variable should appear on the *y*-axis.

Figure 4.14 shows a scatter plot between pretreatment scores and post-treatment scores of the 30 nurses. Note that six cases have the same values and do not appear as separate points. The line is from a fitted straight-line regression of post-treatment on pretreatment score. This would differ from the regression the other way around, of pretreatment on post-treatment score, because it finds the line that minimizes the squared errors or *residuals*. These are measured by the vertical distances between the points and the line using a method of estimation known as *least squares*.

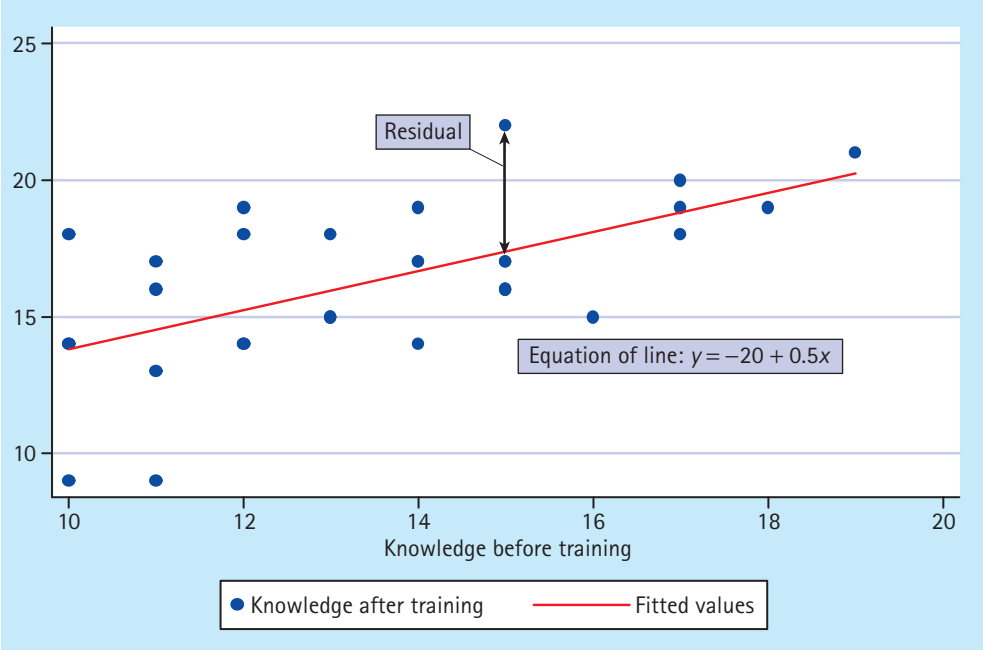

Figure 4.14 Scatter plot showing linear regression line

Regression assumes that one variable (the dependent variable on the *y*-axis) is a random variable whereas the other (the independent or explanatory variable) is a fixed error-free measurement. This is of course rarely the case and is not true of the nurse data. Errors of measurement in the independent variables attenuate the slope of the line (make it less steep), and it is worth bearing this in mind, although in practice this feature is usually ignored.

The difference between regression and correlation

Although mathematically (and in terms of the situations where they are used) correlation and regression are closely allied, there are conceptual differences between them. When one of the variables is thought to depend on or be predicted by the other, the natural approach is that of regression (see below). When the question is simply a matter of the degree of an association, then correlation analysis is more appropriate. Correlation is essentially symmetrical, and distributional assumptions about both variables are necessary for hypothesis tests, whereas in regression no distributional assumptions are made about the independent variables. In the above example, one could treat the data either way: correlation between the scores could be estimated, or a regression of knowledge after training could be fitted to that before training (the reverse relationship would probably not make much sense, indicating the importance of actively choosing what is to be the dependent variable and bearing in mind the criteria for causation, e.g. as in the Bradford-Hill criteria listed earlier).

Pearson's product moment correlation

The most commonly used measure of association is the Pearson *product moment correlation coefficient*, *r*, used to measure the strength of a linear relationship between two normally distributed variables with equal variances. It ranges between −1 (perfect negative relationship) and +1 (perfect positive relationship). The square of *r* is the percentage of variance in one variable explained by the other. In the nurse example, *r* for the relationship between pre- and post-training knowledge is 0.52. Pretraining scores explain therefore 0.52^2 or 27 per cent of the variance of the post-training scores. Hypothesis tests for the population correlation ρ can be based on the *Fisher transformation*:

$$z = 1/2 \ln((1 + r)/(1 - r))$$

which produces a test statistic *z* that is approximately $N(0,1)$. A hypothesis test (single-sample *z*-test) of the null hypothesis that the population correlation is 0 shows that the observed *r* of 0.52 is significantly different from 0 ($P = 0.003$, 95 per cent confidence interval 0.20 to 0.74).

A correlation coefficient of 0 means that there is no linear relationship, but this does not mean that there is no relationship of any kind – it might be non-linear (hence the need to plot the data). It is important that there is no funnel shape in the data where the spread widens or narrows at different levels of the variables (indicating *heteroscedasticity*, i.e. a change in variance) and no outliers, to which Pearson's correlation is very vulnerable. *Partial correlations* are correlations between a pair of variables after adjusting

for a third (and can be derived from the three pair-wise correlations).

Spearman and Kendall's tau correlation

Non-parametric alternatives are available, such as Spearman's rank correlation ρ. This is equivalent to the Pearson's correlation calculated on the ranked data and detects any monotonic (i.e. increasing or decreasing) relationship, such as an exponential relationship. Spearman's ρ ranges from −1 to 1 and is interpreted in a similar way to Pearson's *r*. However, ρ^2 cannot be interpreted as a percentage of variance explained in the same way as r^2. If there are tied observations in the data, a corrected version should be used. Furthermore, if the sample size is small (< 15) or if there are tied observations among the data, either exact *P* values should be computed or a statistical table for critical values of ρ should be consulted.[29] Spearman's ρ assumes that the difference between all subsequent ranks is the same (e.g. the difference between ranks 1 and 2 is the same as between ranks 10 and 11). This a commonly made assumption for Likert scales in psychiatric measurement scales. If the assumption of equidistance cannot be assumed, then Kendall's tau correlation coefficient is an alternative for ordinal-level variables.

Linear regression

Multiple linear regression is used for prediction of a variable from a set of other variables, or to control for confounders when the interest is in one particular variable (e.g. treatment arm). For example, one might use a *t*-test to compare the mean values of a continuous variable, such as symptom score, between two groups in a randomized controlled trial. Then one might wish to control for the baseline value and perhaps also sociodemographic variables using regression. This procedure is often known as ANCOVA. Regression and ANOVA, and ANCOVA (a mixture of the two), are examples of *general linear models* – not to be confused with the even more general *generalized linear models* (see below).

The example below shows typical output for a regression of functioning (GAF score) on days in hospital, randomization group of the patient's nurse and admission cohort (three categories, included as two dummy variables). (Because the patients are 'clustered' within the nurse, a preferable analysis would take this into account; see Multivariate and other more complex techniques, below) (see Stata output, below).

The linear regression prediction equation is (rounded):

GAF score = 59.39 + 7.470 (if randomization group = 1) + 2.844 (if admission group = 2) + 21.952 (if admission group = 3) + 0.00323 (days in hospital)

The coefficients are the estimates of the increase in the dependent variable for a one-unit increase in the independent variable, adjusted for all other variables. Thus, since *group* is coded 0 and 1, the coefficient for *group* in the

```
xi: regress gaf group i.admitgrp days
i.admitgrp        _Iadmitgrp_0-2      (naturally coded; _Iadmitgrp_0 omitted)

      Source |       SS       df       MS              Number of obs =      60
-------------+------------------------------           F(  4,    55) =   12.08
       Model |  4122.47021     4  1030.61755           Prob > F      =  0.0000
    Residual |  4694.11312    55  85.3475113           R-squared     =  0.4676
-------------+------------------------------           Adj R-squared =  0.4289
       Total |  8816.58333    59  149.433616           Root MSE      =  9.2384

------------------------------------------------------------------------------
         gaf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |   7.470304   2.653043     2.82   0.007     2.153487    12.78712
_Iadmitgrp_1 |   2.844533   2.652026     1.07   0.288    -2.470246    8.159311
_Iadmitgrp_2 |   21.95218   3.820844     5.75   0.000     14.29504    29.60933
        days |   .0032465   .0086497     0.38   0.709    -.0140879    .0205808
       _cons |   59.39308   2.743692    21.65   0.000      53.8946    64.89156
------------------------------------------------------------------------------

. testparm  _Iadmitgrp_*

 ( 1)  _Iadmitgrp_1 = 0
 ( 2)  _Iadmitgrp_2 = 0

       F(  2,    55) =   17.15
            Prob > F =    0.0000
```

regression equation is the adjusted difference between groups. With no other variables in the equation, it would be the same as a *t*-test. The R^2 is the percentage of variance explained and is equal to the correlation between the predictions made by the above equation and the dependent variable. R^2 depends on the number of parameters; a value that is more comparable across different numbers of parameters is the adjusted R^2, which takes these into account. The F statistic in the top right corner reflects the significance of the group of explanatory variables taken together. The F statistic at the end reflects the significance of admission cohort as a categorical variable. Specific comparisons between the dummy-coded diagnostic groups 2 and 3 and the reference category appear in the main body of the result: $P = 0.288$ and $P < 0.001$, respectively.

Logistic regression

If one wished to compare the groups in terms of a binary variable, for example symptoms improved, then the approach would be a χ^2 test to compare proportions followed by logistic regression to control for other variables. Logistic regression analysis maximizes the *likelihood* (the probability p of the observed data given a particular statistical model) to estimate the parameters of a linear equation for the log of the odds – $\ln(p/(1-p))$ – also called the *logit*.

It does this through an iterative process to maximize the likelihood (*maximum likelihood estimation*). The output can be expressed in terms of the coefficients that apply to the logit through the linear predictor. An alternative formulation is in terms of exponentiated coefficients, which are the estimated adjusted ORs, and these are usually more useful for interpretation.

The example below shows the output in OR format for a logistic regression of binary (0/1) functioning category against randomization group (coded 0/1) and three-category diagnostic group for 30 of the patients (see Stata output ①, below).

Here, diagnosis 2 seems to be significant at about $P = 0.05$ compared with diagnosis 1. However, we need to test the whole variable, so we still need to perform a likelihood ratio test. We rerun the analysis without diagnosis to see whether the likelihood drops significantly, as illustrated below. Such tests can be validly applied when model A is *nested* in model B (i.e. A's parameters form a subset of B's) (see Stata output ②, below).

① Logistic regression

```
                                                  Number of obs   =         30
                                                  LR chi2(3)      =       7.76
                                                  Prob > chi2     =     0.0512
Log likelihood = -16.913369                       Pseudo R2       =     0.1866

------------------------------------------------------------------------------
     gaf_cat | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |   4.444985   3.788074     1.75   0.080     .8364988    23.61975
_Idiagnosi_2 |   7.141792   7.263598     1.93   0.053     .9729447    52.42353
_Idiagnosi_3 |   1.903212   2.285294     0.54   0.592     .1808843    20.02503
------------------------------------------------------------------------------
```

② estimates store A

```
logistic gaf_cat group

Logistic regression                               Number of obs   =         30
                                                  LR chi2(1)      =       3.40
                                                  Prob > chi2     =     0.0653
Log likelihood = -19.095425                       Pseudo R2       =     0.0817

------------------------------------------------------------------------------
     gaf_cat | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |          4   3.098387     1.79   0.074     .8764398    18.25567
------------------------------------------------------------------------------

. estimates store B

. lrtest A B

Likelihood-ratio test                                 LR chi2(2)  =      4.36
(Assumption: B nested in A)                           Prob > chi2 =    0.1128
```

The likelihood ratio test compares the likelihoods with and without diagnosis and shows that there is no significant reduction. Diagnosis as a whole is not significant and can be omitted from the model.

The output below illustrates the use of the linear predictor format as an alternative to the OR format to obtain logit (or log odds).

This linear predictor can be used to form predictions of the odds and hence the probabilities for individuals, as shown in the example for someone aged 20 years in group 1:

Logit or ln(odds) = 3.466 + 2.132 − 20 × 0.2783 = 0.0323
Odds = exp(0.0323) = 1.0328
Probability = 1.0328/(1 + 1.0328) = 0.5081

The adjusted ORs for a positive outcome (gaf_cat = 1) in relation to a unit increase in the independent variables are found by exponentiating the coefficients, for example exp(2.131) = 4.

Logistic regression is commonly used in case–control studies: the case–control status is the binary dependent variable and the ORs measure the degree of association between the risk or protective factors and that status. In this situation the ORs are often interpreted as RRs, although this is numerically accurate only when the risk is very small (e.g. < 0.05). It is true, however, that, whatever the risk, a significant OR will indicate a significant RR, even if numerically they are not equivalent. Note that the calculation of probabilities as above cannot be performed: this is because the proportion of cases is determined by the design of the study (50% each in 1 : 1 matching), and so calculating a probability based on these proportions is meaningless.

Standardization of coefficients

In a regression analysis, either or both dependent and independent variables can be *standard scored* or *z*-scored – that is, the mean is subtracted and the value divided by the sample standard deviation (cf. standard effect sizes). The resulting coefficients are sometimes called 'beta coefficients'. If both dependent and independent variables are standardized, then the betas have a similar interpretation to partial correlation coefficients, although they are numerically slightly different. This is a useful way of comparing (within a given regression model) the relative sizes of coefficients for variables that are measured in different units. For logistic regression, one would only standardize the independent variables, because the dependent variable, being binary, does not need to be standardized.

Conditional logistic regression

One further type of regression that is especially important for individually matched case–control studies and other situations where the observations are not independent is *conditional logistic regression*. Here, the observations that are individually matched (e.g. in a case–control study) or connected in some other way (e.g. the two patients of each nurse in the example) are indicated by a 'link' variable or 'match group' (e.g. the nurse identification in the example). The likelihood that is maximized in estimating the parameters of a logistic model for case status is conditional on membership of the linked group and so effectively controls for characteristics that the members of the linked group have in common. This is the point of matching – to control for unmeasured confounders – and so it is important to make use of conditional logistic regression rather than standard logistic regression where appropriate, otherwise the benefits of matching may be wasted.

Other types of regression

Some other types of regression worth knowing about are *multinomial logistic regression*, where the dependent variable is categorical non-ordered, such as diagnosis; *ordered logistic regression*, where the dependent variable is ordered categorical, such as grade of illness; and *Poisson regression*, for count data, such as the number of admissions to hospital, or binary data, such as admitted or not where the time at risk is either measured or implied and one is interested in rates rather than proportions. Poisson regression is commonly used in cohort studies to estimate rate ratios over a follow-up period after controlling for confounders and produces *rate ratios*. *Cox regression* is appropriate where the outcome is time to an event, as in survival analysis. It produces *hazard ratios*, which are ratios of the risks of an event at any given point in time (rather than averaged over the follow-up time, as in Poisson regression). A more general model, which encompasses all of the above as well as linear and logistic regression and a number of others, is the *generalized linear model*.

Model-building in regression

The best way to include variables in a regression model of any type is on the basis of theoretical background in the subject matter of interest. However, sometimes investigators are faced with a large number of potential independent variables without much guidance as to which to use. A rule

gaf_cat	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
group	2.13168	.9555024	2.23	0.026	.2589298	4.004431
age	-.2782977	.157653	-1.77	0.078	-.5872919	.0306965
_cons	3.466657	2.40287	1.44	0.149	-1.242881	8.176196

of thumb is to have at least ten cases per parameter to be estimated, so there is often a need to cut down the number of variables included. Some possible automatic methods are as follows:

- *Backward selection:* first include all variables, then omit variables one by one, starting with the least significant and refitting the model each time.
- *Forward selection:* start with the single most significant variable (having tested each separately), then progressively add in a variable at a time (again refitting the model each time).
- *Stepwise:* a mixture of the above procedures.

None of these is necessarily optimal, and they may give rise to different sets of variables. See Derksen and Keselman[36] for a discussion. An increasingly popular alternative to model selection based on significance testing is the use of information criteria, such as the Akaike information criterion (AIC).[37] This allows the comparison of the likelihood of a set of different candidate models, which need not be nested, as is the case with likelihood ratio tests.

MULTIVARIATE AND OTHER MORE COMPLEX TECHNIQUES

Principal components and factor analysis

Principal components analysis (PCA) transforms the data variables into components (linear combinations of original variables) that explain decreasing proportions of the variance in the data and that are uncorrelated.[38] It is essentially a data-reduction technique – often the first few components are plotted to illustrate the main features of multivariate data. The last few are useful for indicating any outliers. It is also used as the mathematical first step in a factor analysis. However, although factor and principal components analysis are often confused, they are conceptually distinct.

Exploratory factor analysis seeks underlying factors that explain the correlation or covariance between variables. The unmeasured variables are assumed to be linear functions of underlying unmeasured or latent variables, plus measurement errors (called residuals or *specific factors*). The analysis is often initiated by running a PCA to determine the number of factors. Factors are extracted using e.g. maximum likelihood methods, which can be rotated, either orthogonally (e.g. using the *varimax* method – axes are kept orthogonal/perpendicular) or obliquely (e.g. using the *oblimin* method – axes are allowed to form oblique patterns). The rotation often allows the coefficients that define the factors (factor loadings) to be interpreted more easily. It is common practice to set loadings to zero if they are below a cut-off (e.g. 0.40), with the aim of finding factors that are each associated with a mutually exclusive set of variables. (The others are not zero, but this presentation simplifies the results.) Costello and Osborne[39] provide some guidance on this method.

Table 4.14 shows typical factor solutions for hypothetical data from a quality-of-life scale. Interpretations of the factors are given along the top (these are the prerogative of the investigator). The percentage of the variance explained by each factor is given along the bottom. In this case the three-factor solution seems to be more interpretable because the variables load into different factors. Note that the negative coefficients do not matter for interpretation (so long as they are consistent within a factor). If scales were to be derived from the factors, then common practice would be to add up the values of the items within factors (i.e. set the factor loadings to 1). It is important that item response scores are in the same conceptual order.

In a *confirmatory factor analysis*, the researcher sets up

Table 4.14 Example of a factor analysis: oblimin-rotated factors based on principal components analysis of eight variables; factor loadings > 0.40 shown

Constituent scales of the SF-36	Two-factor solution		Three-factor solution		
	1 General health	2 Physical health	1 Mental health	2 Physical health	3 Roles
Physical functioning		0.85		0.78	
Role physical	0.41	0.40			–0.83
Bodily pain		0.85		0.82	
General health	0.64		0.70		
Vitality	0.78		0.82		
Social functioning	0.63		0.42		
Role emotional	0.74				–0.72
Mental health	0.91		0.90		
Percentage of variance explained	77%	10%	72%	15%	10%

in advance a hypothesis for the latent constructs and measures a number of variables that reflect different aspects of these constructs. The set of relationships between the observed variables and the constructs is called the *measurement model*, which is then tested for fit against real data. In more complicated models, the factors themselves will have a relationship with one another; this combination of factor analysis and regression analysis is called a *structural equation* and is analysed with specialist software. Figure 4.15 shows how such a theoretical model might be set up. Such diagrams are known as *path diagrams*.

Other multivariate methods

Principal components and factor analysis are widely used methods. Here, we mention some other methods that are much less popular but worth noting. *Discriminant analysis* is a method for classifying individuals into predefined groups by producing linear discriminant functions of the variables (if equal variances can be assumed) or quadratic functions (if not). It also provides a plot in a lower dimension than the original number of variables (analogous to principal components, except that the plot maximizes the variance between group means rather than between individual data points). It has a similar function to binary logistic regression (or multinomial logistic regression when there are more than two groups), but unlike those methods it assumes multivariate normality of the variables used and is hence somewhat more restrictive in application. The main advantage is the plot it produces. Another classification method is *classification and regression tree analysis* (CART); this focuses on finding interactive effects (i.e. combinations of variables) rather than linear functions, as in discriminant analysis or logistic regression analysis, and produces a tree-like diagram (e.g. see Thomas *et al.*[40]).

Correspondence analysis is a method for graphically displaying cross-tabulated data; although it is little used in psychiatry, it is potentially very useful for exploratory

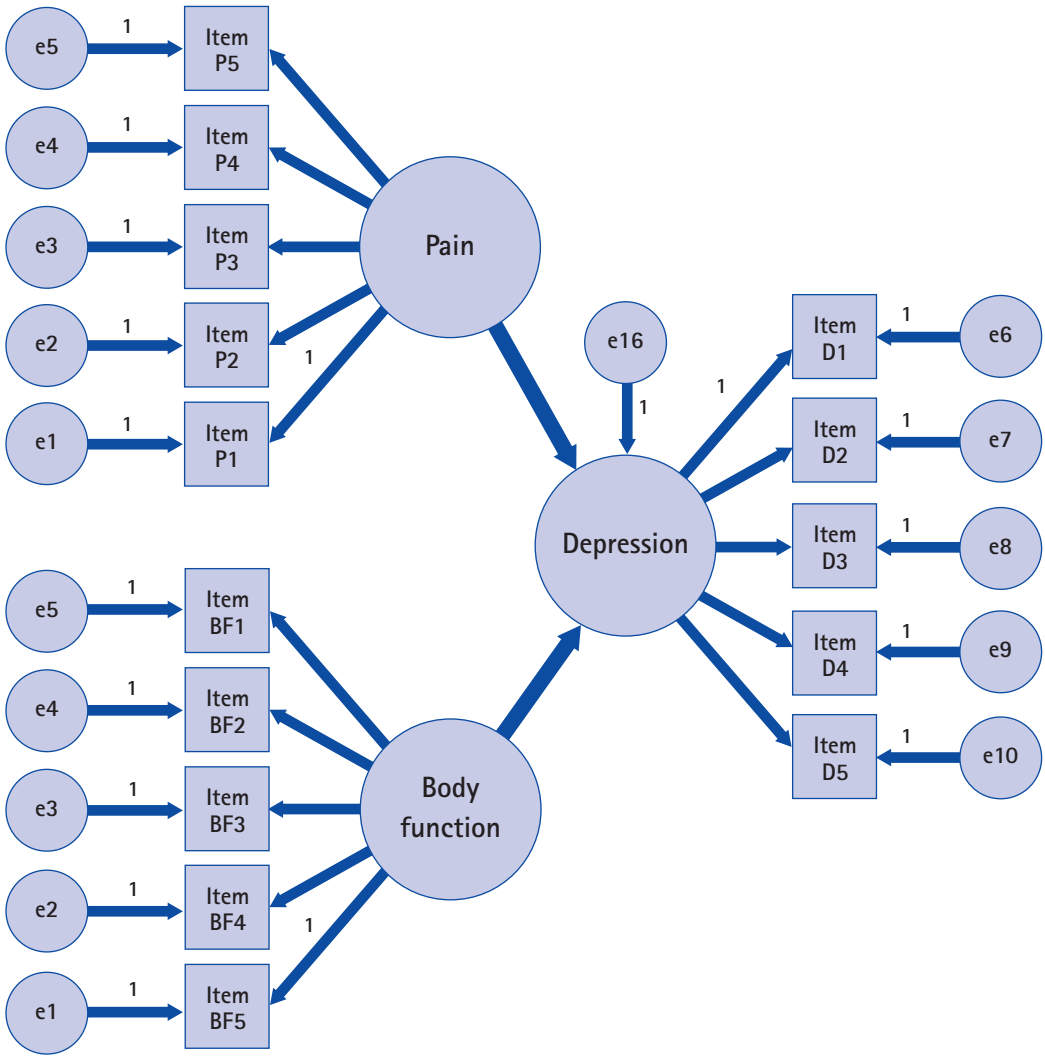

Figure 4.15 Example of a structural equation model. This relates the two latent traits 'Pain' and 'Body function' to the latent trait 'Depression'. Each latent trait is measured by five items (measurement model)

analysis of large cross-tabulations. *Canonical correlation analysis* again is little used, probably because the results are difficult to interpret; it is a method for investigating linear functions that maximize the correlation between the variables in one set with the variables in another set. *Cluster analysis* is a vast set of methods that seek subgroups within a dataset. It is a data-driven exploratory method, overlapping with data-mining, neural networks and pattern recognition. Everitt *et al*.[41] give an overview of the more standard methods. They differ from the classification methods mentioned above in that the groups are unknown a priori – the aim is to use the data to find the groups.

For an introduction to applied multivariate data analysis, see Everitt and Dunn[42] or Tabachnick and Fidell.[43]

Methods for non-independent data

Many standard statistical methods assume that each data point has been sampled independently of the others. However, in some situations this is not the case. Clustered trials and matched case–control studies have already been mentioned. Other examples are sampling of families where all members are interviewed, and of students within classes within schools. Repeated measurements for the same subject – sets of measurements of the same variable at different time points or sets of different variables that measure the same underlying within-subject factor – are also dependent.

In a *multivariate analysis of variance* (MANOVA), the purpose is to compare groups (e.g. gender or experimental condition) in terms of sets of dependent measures simultaneously. It is sometimes known as *profile analysis* when different variables are involved. It has been superseded to some extent by a number of specialized techniques appropriate for such data, such as *multilevel models* and *general estimation equation* (GEE) models. Other terms that will be encountered are *random effects* and *mixed models*. This collection of techniques is very flexible and can cope with within-subject repeated measures and multilevel sampling of different subjects. It can allow the inclusion of several levels of cluster (e.g. individuals clustered within families, families clustered within a therapist, and even cross-clustering, where individuals are clustered within families but family members do not have the same therapist). These modelling techniques allow the inclusion of both individual-level and cluster-level covariates.

To illustrate these, below is a simple regression of GAF score of the 60 patients on randomization group of the nurse in the training example. Because of the clustering of the pairs of patients within the nurses, this would not be correct and would underestimate the standard error for group and hence overestimate the significance.

Overleaf are the same data analysed as a *random effects regression*. Here, the effect of nurse that is common to his or her pair of patients has been explicitly modelled. The standard error increases and the estimate of the group effect is changed slightly, from 8.86 to 7.93 (see Stata output ①, overleaf).

Another approach is to use standard techniques such as linear regression but with specially adapted estimation methods that are *robust* to non-independence (such as the sandwich estimator mentioned earlier). Overleaf are the results for the above example: the point estimate for group is the same as the simple regression, but the standard error has been adjusted to take account of the clustering and is now called 'robust standard error' (see Stata output ②, overleaf).

A simple alternative approach for clustered data is to construct a *summary statistic* for each cluster and then analyse these summary values, for example with a weighted *t*-test using the numbers of subjects as a weight. However, this latter approach is not efficient and does not allow a more complex analysis including individual covariates.

Meta-analysis

Sample sizes in psychiatric studies are sometimes too small to detect clinically important effects. However, it is possible to combine the results of several similar studies by means of meta-analysis.[44,45] A typical meta-analysis starts with a

```
. regress  gaf group

      Source |       SS       df       MS              Number of obs =      60
-------------+------------------------------           F(  1,    58) =    8.90
       Model |  1172.68601     1  1172.68601            Prob > F      =  0.0042
    Residual |  7643.89732    58  131.791333            R-squared     =  0.1330
-------------+------------------------------           Adj R-squared =  0.1181
       Total |  8816.58333    59  149.433616            Root MSE      =   11.48

------------------------------------------------------------------------------
         gaf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |   8.861607   2.970743     2.98   0.004     2.915019    14.80819
       _cons |   63.78125   2.029404    31.43   0.000     59.71896    67.84354
------------------------------------------------------------------------------
```

① `xtreg gaf group,i(nurse)`

```
Random-effects GLS regression                   Number of obs      =        60
Group variable: nurse                           Number of groups   =        30

R-sq:  within  = 0.0154                         Obs per group: min =         2
       between = 0.1563                                        avg =       2.0
       overall = 0.1330                                        max =         2

Random effects u_i ~ Gaussian                   Wald chi2(1)       =      5.34
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0208

------------------------------------------------------------------------------
         gaf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |    7.92753   3.430114     2.31   0.021      1.20463    14.65043
       _cons |   64.21715   2.507726    25.61   0.000      59.3021    69.13221
-------------+----------------------------------------------------------------
     sigma_u |  9.5259075
     sigma_e |  6.6779359
         rho |   .6704923   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

② `regress gaf group, cluster(nurse)`

```
Linear regression                               Number of obs =        60
                                                F(  1,    29) =      5.53
                                                Prob > F      =    0.0257
                                                R-squared     =    0.1330
                                                Root MSE      =     11.48

                                   (Std. Err. adjusted for 30 clusters in nurse)
------------------------------------------------------------------------------
             |               Robust
         gaf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |   8.861607   3.769132     2.35   0.026      1.152867    16.57035
       _cons |   63.78125   2.936454    21.72   0.000      57.77553    69.78697
------------------------------------------------------------------------------
```

systematic review: studies of the same or similar treatments of generally randomized controlled trials are collected in a formalized process.[46,47] Meta-analysis is a statistical technique that allows the combination of the effect sizes extracted from different studies into one overall estimate with an associated confidence interval and a hypothesis test of treatment effectiveness.

The statistical models used in meta-analysis weight the individual study results using either *fixed* or *random-effects* models, the choice between them depending on the degree of between-study variability, which can be tested in relation to pooled variation within studies with a χ^2 test for homogeneity (sometimes called the Q test). If they are homogeneous, then a fixed effects model should be used. In this case, the weights are proportional to the inverse of the within-study variation (e.g. as reported in the standard errors of the effects). If the studies are heterogeneous, then a random-effects model is more appropriate; here, the weight combines the studies' individual precisions with the between-study variation.

Figure 4.16 shows a so-called *forest* plot, a graphical summary of a meta-analysis. This shows the effect sizes obtained from the individual trials with their confidence intervals (the lengths of the lines); the size of the boxes is proportional to the sample sizes, and a vertical line indicates the overall effect.[48]

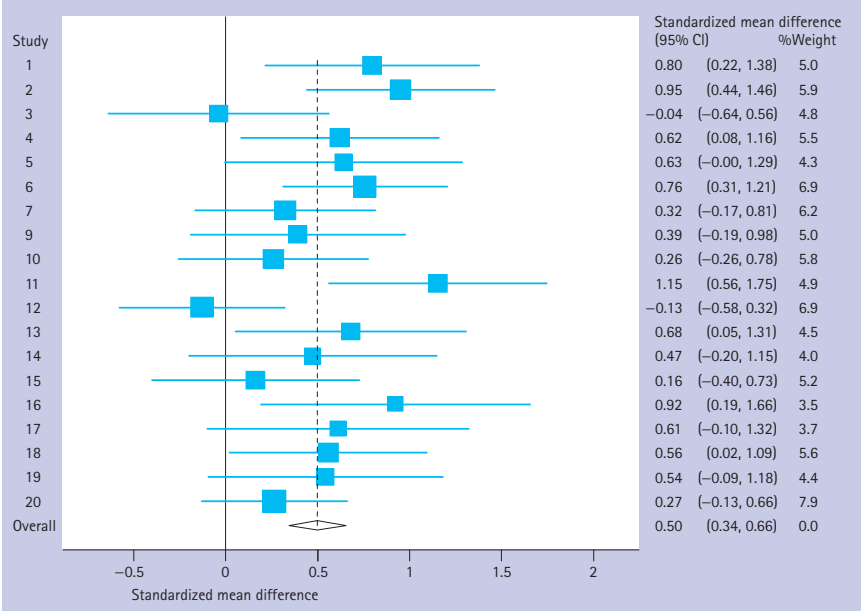

Figure 4.16 Forest plot of a random-effects meta-analysis of 20 studies of a particular type of group therapy. CI, confidence interval.

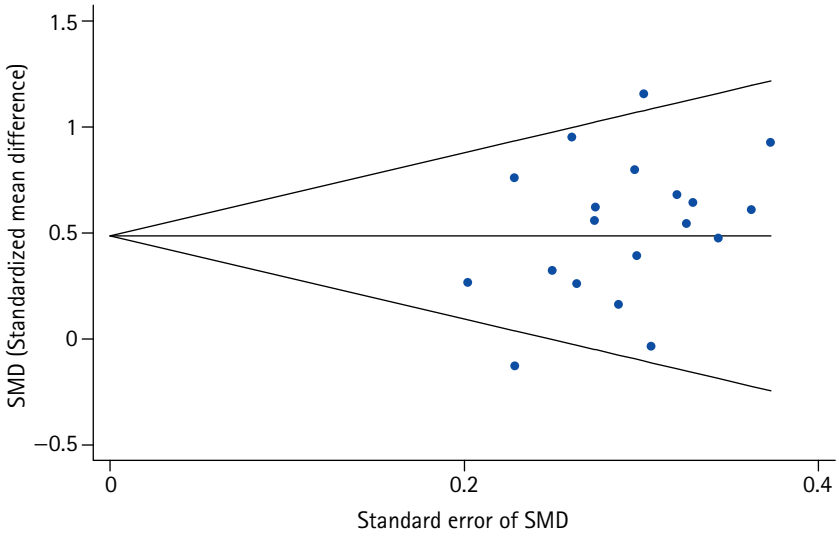

Figure 4.17 Funnel plot for the studies shown in Figure 4.16

A common problem is *publication bias*. Scientists tend to submit only studies with significant results, and editors of journals are more likely to publish such studies. Smaller studies without statistically significant effects or even contradictory effects are less likely to be published, so that the overall effect tends to be overestimated. The possibility of bias can be assessed informally by a *funnel plot* (Figure 4.17). This is a simple scatter plot of the sample sizes (or precision) of the studies against their estimated effect sizes.

The funnel plot in Figure 4.17 shows such a gap (bottom right), which suggests a publication bias although a formal test for bias (Begg's adjusted rank test) was not significant.

KEY POINTS

- Validity types include face; content; criterion; concurrent; discriminant; predictive; construct. Test–retest and inter-rater reliability are assessed by kappa coefficient (categorical data), intraclass correlation (continuous data), and item consistency by Cronbach's alpha.
- In diagnostic testing, post-test odds = pre-test odds × likelihood ratio where odds = probability/(1 − probability); receiver operator curves, sensitivity, specificity (both independent of base rate), PPV and NPV are also key concepts.
- The randomized controlled trial is the gold standard study design, and involves the following concepts: phases of development,

equipoise, placebo controls, blinding, intention-to-treat analysis, individual v. cluster randomization and the Consort diagram.

- Epidemiological designs include case–control studies (which are analysed using odds ratios, often via logistic regression), and cohort studies (which may use rates – or 'person-year analysis'– rate ratios and Poisson regression).
- Partly because of the central limit theorem, many statistical techniques are based on the bell-shaped normal or Gaussian distribution, characterized by the mean and standard deviation (square root of variance); the mean $\pm$ 1.96 $\times$ SD contains 95% of the observations.
- The approximate (large sample) 95% confidence interval for an estimate is the estimate $\pm$ 1.96 $\times$ its standard error, and indicates its uncertainty; the size of the estimate compared to its standard error produces the *P* value, the chance of obtaining so extreme a value under the null hypothesis (if *P* is below a critical value, alpha, it is said to be significant at that level, e.g. 0.05).
- Key basic techniques are: *t*-test (non-parametric Mann–Whitney or equivalently the Wilcoxon rank sum) and paired *t*-test (non-parametric Wilcoxon signed rank); chi-square test and McNemar test for proportions; analysis of variance, correlation (Pearson and non-parametric Spearman) and regression (linear for continuous and logistic for binary outcomes), and survival analysis (Kaplan–Meier, log-rank and Cox regression).
- Important multivariate techniques include principal components analysis, factor analysis (exploratory and confirmatory), discriminant analysis, cluster analysis, structural equation modelling and path analysis.
- Meta-analysis allows one to combine the results of several similar studies into one overall estimate with an associated confidence interval, and uses forest plots (to show the estimates) and funnel plots (to detect possible publication bias).

REFERENCES

1. Bland JM (2000) *An Introduction to Medical Statistics*. Oxford: Oxford University Press.
2. Petrie A, Sabin C (2005) *Medical Statistics at a Glance*, 2nd edn. Oxford: Blackwell.
3. Everitt BS (2006) *Medical Statistics from A to Z*, 2nd edn. Cambridge: Cambridge University Press.
4. Dupont WD (2002) *Statistical Modelling for Biomedical Research: A Simple Introduction to the Analysis of Complex Data*. Cambridge: Cambridge University Press.
5. Beck AT, Guth D, Steer RA, Ball R (1997) Screening for major depression disorders in medical inpatients with the Beck Depression Inventory for Primary Care. *Behavior Research and Therapy* **35**: 785–91.
6. Irvine SH, Kyllonen PC (eds) (2002) Mahwah, NJ: *Item Generation for Test Development*. Lawrence Erlbaum Associates.
7. Streiner DL, Norman GR (2003) *Health Measurement Scales: A Practical Guide to their Development and Use*, 3rd edn. Oxford: Oxford University Press.
8. Shrout P, Fleiss JL (1979) Intraclass correlation: uses in assessing rater reliability. *Psychological Bulletin* **86**: 420–28.
9. Bland J, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **i**: 307–10.
10. Brennan RL (2001) *Generalizability Theory*. New York: Springer-Verlag.
11. Pocock SJ. (1994) *Clinical Trials: A Practical Approach*. Chichester: John Wiley & Sons.
12. Moher D, Schulz KF, Altman DG, CONSORT Group (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* **357**: 1191–4.
13. Moher D, Schulz KF, Altman DG (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* **357**: 1191–4.
14. Campbell MK, Elbourne DR, Altman DG, *et al.* (2004) CONSORT statement: extension to cluster randomised trials. *British Medical Journal* **328**: 702–8.
15. Donner A, Klar NS (2000) *Design and Analysis of Cluster Randomisation Trials in Health Research*. London: Arnold.
16. Deeks JJ, Dinnes J, D'Amico R, *et al.* (2003) Evaluating non-randomised intervention studies. *Health Technology Assessment* **7**: 1–173.
17. Morton V, Torgerson DJ (2003) Effect of regression to the mean on decision making in health care. *British Medical Journal* **326**: 1083–4.
18. Sackett D, Haynes RB, Tugwell P, Guyatt GH (1991) *Clinical Epidemiology: A Basic Science for Clinical Medicine*, 2nd edn. Boston, MA: Little Brown.
19. Phelps CE (2002) *Health Economics*, 3rd edn. Boston, MA: Addison-Wesley.
20. Sackett DL (1979) Bias in analytic research. *Journal of Chronic Diseases* **32**: 51–63.
21. Lee W, Bindman J, Ford T, *et al.* (2007) Bias in psychiatric case–control studies: a systematic review of current practice. *British Journal of Psychiatry* **190**: 204–9.
22. Hill AB (1965) The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine* **58**: 295–300.
23. Höfler M (2005) The Bradford Hill considerations on causality: a counterfactual perspective. *Emerging Themes in Epidemiology* **2**: 11.
24. Miller JN (1993) Outliers in experimental data and their treatment. *Analyst* **118**: 455–61.
25. Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates.
26. Fitzpatrick R, Davey C, Buxton MJ, Jones DR (1998)

Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment* **2**: 1–74.
27. Shaffer JP (1995) Multiple hypothesis testing. *Annual Review of Psychology* **46**: 561–84.
28. Howell DC (1997) *Statistical Methods for Psychology*, 4th edn. Belmont, CA: Duxbury.
29. Siegel S, Castellan NJ (1988) *Nonparametric Statistics for the Behavioural Sciences*, 2nd edn. New York: McGraw-Hill.
30. Altman D, Bland NJ (2004) Confidence intervals illuminate absence of evidence. *British Medical Journal* **328**: 1016–17.
31. Bolstad WM (2007) *Introduction to Bayesian Statistics*, 2nd edn. New York: John Wiley & Sons.
32. Spiegelhalter DJ, Abrams KR, Myles JP (2003) *Bayesian Approaches to Clinical Trials and HealthCare Evaluation*. Chichester: John Wiley & Sons.
33. Everitt BS (1992) *The Analysis of Contingency Tables*. London: Chapman & Hall
34. Kleinbaum D, Klein M (2005) *Survival Analysis: A Self-Learning Text*. New York: Springer.
35. Collet D (1994) *Modelling Survival Data in Medical Research*. Boca Raton, FL: Chapman & Hall/CRC.
36. Derksen S, Keselman HJ (1992) Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology* **45**: 265–82.
37. Anderson DR (2008) *Model Based Inference in the Life Sciences: A Primer on Evidence*. New York: Springer.
38. Joliffe IT (1992) Principal component analysis and factor analysis. *Statistical Methods in Medical Research* **1**: 69–95.
39. Costello AB, Osborne JW (2005) Best practices in exploratory factor analysis: four recommendations for getting the most from your factor analysis. *Practical Assessment, Research and Evaluation* **5**: 1–9.
40. Thomas S, Leese M, Walsh E, *et al*. (2005) A comparison of statistical models in predicting violence in psychotic illness. *Comprehensive Psychiatry* **46**: 296–303.
41. Everrit B, Landau S, Leese M (2001) *Cluster Analysis*, 4th edn. London: Edward Arnold.
42. Everitt BS, Dunn G (2001) *Applied Multivariate Data Analysis*. London: Edward Arnold.
43. Tabachnick BG, Fidell LS (2007) *Using Multivariate Statistics*, 5th edn. Boston, MA: Allyn & Bacon.
44. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F (2002) *Methods for Meta-Analysis in Medical Research*. New York: John Wiley & Sons.
45. Rosenthal R (1991) *Meta-Analytic Procedures for Social Research*. Newbury Park, CA: Sage.
46. Davies HTO, Crombie IK (2001) What is a systematic review? *Bandolier* **1**: 1–6.
47. Egger M, Davey Smith G, Altman DG (2003) *Systematic Reviews in Health Care: Meta-Analysis in Context*. London: BMJ Books.
48. Lewis S, Clarke M (2001) Forest plots: trying to see the wood and the trees. *British Medical Journal* **322**: 1479–80.

Chapter 5

Epidemiology

Stephen Stansfeld and Charlotte Clark

INTRODUCTION

Epidemiology is the study of health in relation to populations. Succinctly, it has been described as the basic quantitative science of public health. It differs from clinical medicine in that the results of epidemiological research have relevance to public health with consequences for health at the level of the population rather than the individual. It has been described as the study of the distribution and determinants of health-related states or events in specified populations and the application of this to the control of health problems.

Epidemiological methods have a number of useful functions: to describe the burden of illness in a population; to elucidate the aetiology of disease and biological dysfunction; to test the effectiveness of health interventions; and to assess need for services.

Historically, psychiatric epidemiology began to develop from purely descriptive to analytical epidemiology, aimed at understanding the aetiology of disorders in the mid-1970s.[1] This began with empirical work on improving the reliability of psychiatric diagnoses, shown in the 1960s to be very unreliable. In the US/UK comparative diagnostic study, rates of schizophrenia in New York and London, which were thought to be very different, were found to be similar when standard methods of diagnosis were used in both settings. Operational criteria for a range of disorders were then developed and standardized psychiatric interviews were used to identify specific psychiatric disorders. Diagnoses such as schizophrenia were also strengthened by genetic adopted-away studies, which confirmed genetic rather than solely cultural transmission of disease between generations.[2] An important study of the time was the World Health Organization (WHO) International Pilot Study of Schizophrenia (IPSS), which found that a core syndrome of schizophrenia could be diagnosed reliably in different settings across the world by using a standardized psychiatric interview. More recent studies have questioned whether schizophrenic syndromes are equally frequent across cultures, but at the time this study greatly increased confidence in pursuing cross-cultural epidemiological research. This led to developing new diagnostic instruments that could be administered by interviewers without clinical training in large epidemiological field studies, such as the Epidemiologic Catchment Area Study (ECA) in the 1980s. This was followed in the 1990s by the National Co-morbidity Survey (NCS) and its replication study (NCS-R). Genetic epidemiology has progressed from family, twin and adoption designs to linkage including whole-genome studies. Analytical epidemiology has made substantial progress, for instance identifying, for schizophrenia, genetic inheritance, obstetric complications and, intriguingly, urban and winter birth as risk factors.

The advantages of studying an illness in a population rather than a clinical sample are several. Clinical samples, especially those seen in psychiatric out-patient and in-patient settings, are often the more severe or complex cases. Epidemiological samples give a better idea of distribution of the full range of illness severity, including milder cases, that may be helpful for understanding disease aetiology, and results found in epidemiological samples can be applied to the population from which they were taken. The type of bias observed through the choice of a sample of patients recruited from a health services source, to do with the selection processes involved in referral to health services, is called Berkson's bias.

Epidemiological studies are often used in determining causation of disease, where the potential aetiological or causal variables are referred to as the 'exposure', to be related to the dependent variable, the illness 'outcome'.

RATES OF DISEASE

The distribution of illness in a population is expressed in terms of rates of disease, which allow the frequency of diseases to be compared within and across populations (Box 5.1). There are two important rates: the prevalence rate and the incidence rate. Prevalence rate assesses the proportion of a specified illness or condition in a defined population. Point prevalence refers to the prevalence of illness at a specific point in time (e.g. on a census day), whereas period prevalence refers to the number of cases counted during a longer period (e.g. 6 months).

Prevalence is distinguished from incidence because incidence rates include the time dimension (see Box 5.1).

Box 5.1 Rates of disease

- *Rate of disease:* frequency of state or event/number of population at risk
- *Prevalence rate:* proportion of a defined population having a condition at any one time
- *Incidence rate:* proportion of a defined population developing a condition within a stated period.

Hence, incidence rates deal with the number of new cases or episodes of illness developing in a population over a unit time, usually a year. Strictly, for entirely new cases of an illness, the term 'inception rate' rather than 'incidence rate' is correct because incidence may include recurrent episodes.

Rates are always expressed in relation to the population at risk as the denominator. In other words, the denominator is specified as the population that is potentially at risk of developing the illness. For example, prevalence rates of Alzheimer's disease will be expressed as a proportion of the elderly population – it would make no sense to include children in the at-risk population.

Types of study in epidemiology

Epidemiological data can be collected either through new studies or by using routine statistical data. Useful routine data include statistics on mortality and cause of death drawn from death certificates, which can be accessed in the UK from the National Health Service (NHS Central) Registry and have nearly complete data on deaths. Other routine sources of data include the national cancer registry and birth certificates. In some areas, detailed systematic information is collected on health service attendees, including diagnoses, as a case register. When these data are complete and relate to a specific geographical area, they can be very useful for assessing burden of disease and trends in incidence of diagnosed disease. However, routine data sources should be treated cautiously, as there are often missing data. Usually, in order to understand the aetiology of a psychiatric disorder, it is necessary to carry out an epidemiological study. Broadly, studies may be either cross-sectional or longitudinal. Cross-sectional studies include straightforward prevalence studies and analytical case–control studies. Longitudinal studies include incidence and cohort studies.

Cross-sectional studies

A cross-sectional study is a study in which any association between two factors relates to a single point in time. Their uses include estimating needs for health services, providing statistics for health education, and screening for undiagnosed disease. Cross-sectional studies are quicker and cheaper than longitudinal studies, can be applied to the relevant population, and investigate hypotheses that are relevant at that point in time. Their disadvantages are that it is difficult to interpret associations in terms of causation, they are not suitable for rare or short-duration diseases, and they provide no estimate of incidence. They can, however, provide prevalence estimates, although these may be biased by movers in or out of the study area.

Case–control studies

In case–control studies, a type of cross-sectional study, a representative group of cases is compared with a group without disease that is representative of the population from which the cases derive. Unlike in observational cross-sectional studies, subjects are sampled on disease rather than exposure. All possible eligible cases should have an equal likelihood of selection; for example, patients with depression who undergo cognitive-behavioural therapy (CBT) could be matched to a control group of patients with depression who do not undergo CBT. Controls are then matched to cases either on an individual or a group basis. Matching is typically carried out for age, sex and sometimes social class, depending on the study. If subjects are inadvertently matched on a factor that is part of the study question, this is referred to as 'overmatching'.

As the selection of subjects is based on cases, this study design is useful for rare diseases, can involve screening for a wide range of possible risk factors and expensive or time-consuming tests, can test current hypotheses, and is not biased by dropout over time. This type of study is often seen as 'quick and dirty', as there is scope for bias in the selection of cases and controls, measurement bias and recall bias (see below for discussion of bias).

Longitudinal studies

In longitudinal studies, also referred to as prospective studies, a group of people are studied over a period of time. Certain characteristics are determined at the beginning of the study, and the incidence of disease in the group is then observed. This is a classical incidence study. Alternatively, in a cohort study, a group of subjects is studied at baseline and then followed up with recurrent waves of data collection, which collects data on new exposures and outcomes. Examples are the British Birth Cohort Studies that follow up a sample born in a week in 1946, 1958, 1970 and 2000.[3,4] A specialized form of cohort studies are occupational cohorts, where a group employed in the same organization is followed up over time (e.g. Whitehall II Study of civil servants[5]) or populations with special exposures are followed (e.g. victims of a disaster followed up for post-traumatic stress disorder[6]). The advantage of longitudinal studies over cross-sectional studies is that incidence rate can be measured, the longitudinal association of exposures and outcomes is more likely to indicate causation, and the longitudinal observation of cases gives a truer picture of

severity, fatality, impact of therapy, and social class distribution of disease. Importantly, risk factor assessment at baseline is unbiased by presence of disease that develops by follow-up. Additionally, information about changing risk factor states may be obtained, and information given by subjects is not open to bias. However, these studies are expensive, gathering of results is slow, and refusals and dropouts may introduce bias. It is also difficult to maintain constant measurement techniques over time with changes in study researchers. In incidence studies, there is a need to test fixed hypotheses, and this design is not suitable for rare diseases.

Clinical trials

The intervention study or clinical trial can provide higher-quality data than observational studies because the investigator randomly allocates the exposure status (treatment) of the participants, thus ensuring greater validity to the results. Participants are randomly allocated to 'treatment' or 'control' groups. It is slightly more complex than that because intervention studies may be therapeutic (determining the ability of a treatment to diminish symptoms, prevent recurrence or decrease mortality) or preventive (testing whether a procedure can reduce the risk of developing a disease). Subjects are selected from an experimental population in which the trial is to be carried out; this will be a subset of the reference population to whom the results of the trial would be expected to apply. It is essential to select a large enough sample to give sufficient power to test the hypotheses and a sample that has enough outcomes (e.g. recovery from depression) to allow meaningful comparisons between the study groups within the study follow-up period.

Study subjects are invited to take part, having been informed of the nature of the study and the fact that they may be allocated to a control group without the active treatment. Potential participants are then screened to ensure they fit the inclusion criteria or may be ruled out on the basis of predefined exclusion criteria; such criteria control the population sampled in the study. Randomization of subjects to intervention or control group should be carried out in a double-blind fashion, so that neither the subject nor the investigator is aware of which group they are in. This reduces bias related to the subject's report and bias in the investigator's assessment of outcomes that could be influenced by knowledge of whether the subject was in the treatment or control groups (observation bias). This might also influence allocation to treatment or control group (e.g. investigators might tend to allocate more severe cases to the treatment rather than the control group). Another advantage of randomization is that groups will be similar in background and both known and unknown confounding variables.

A type of non-randomized intervention study uses a historical control group in which a group of patients allocated a new treatment is compared with a group that received an earlier, less advanced treatment. If the new treatment is much more effective than the old treatment, then a difference may be observed. However, if the difference between the groups is small, then it may be difficult to be certain whether the new treatment is more effective because of confounding by changes over time in patient populations, diagnoses or other treatments.

Compliance is an important issue, as non-compliance with, for example, a drug treatment will weaken the difference between treatment and control groups. Similarly, it is important to minimize dropout from the trial; frequent contact between investigators and participants can help to avoid this. To reduce bias in analysis, all subjects must be included (not excluding dropouts) in what is called 'intention to treat' analysis – 'once randomized, always analysed' – otherwise there may be a temptation to exclude subjects with less favourable outcomes.[7]

The progress of epidemiological knowledge is not straightforwardly linear. Often results of studies on the same topic are contradictory or show varying strengths of association between risk factors and outcomes. Observational studies may be subject to unmeasured confounding factors and may be contradicted by the results of randomized controlled trials (RCTs). How should the results of such a bewildering array of scientific knowledge be synthesized? Systematic reviews of the literature attempt to bring together all relevant studies on a topic by scanning the range of relevant databases (e.g. PubMed, PsycINFO, Web of Science), including those papers that fulfil predetermined quality criteria and then summarizing the results narratively or statistically. Where studies have comparable estimates of the magnitude of effects, these can be combined to give a summary measure of size of effect through a technique called meta-analysis. Pooled estimates of magnitude of effects are extremely useful as summary measures and are usually accompanied by measures of heterogeneity showing how consistent these effects are across studies and subgroups, such as by gender or age.

Selecting a sample

In epidemiological studies, it is important to select a sample that is representative of the population you are studying. The simplest method is to use random sampling to randomly select a sample from a defined population, so that all individuals within the population of interest should have an equal chance of being surveyed. This, however, is often difficult to achieve; the approach usually taken instead is to define a population that is thought to be representative of the population of interest and to sample from within this population. Bias can occur in the selection of the population fairly easily; for example, people who have an interest in the topic you are researching are more likely to participate in the study; alternatively, the use of a Web-based survey to collect data rules out the participation of people who do not have access to a computer. If you need greater

numbers of one population group in your sample, then you may want to oversample in certain groups (e.g. women or elderly people) and you may use stratified sampling with a higher proportion or 'sampling fraction' from the specific groups. Samples are usually selected from a convenient source of names that covers the relevant population – a sampling frame – such as the electoral roll, general practice age–sex registers and postcode address files.

Bias

There are several important types of bias or potential error in epidemiological studies that you need to be aware of because they may lead to misleading results. In *selection bias*, errors arise from the initial identification of the study population – for instance, where the rate of response to the study is related to exposure status; this may be particularly an issue when control subjects are recruited from groups of students or staff or from newspaper advertisements.[8] *Observation (information) bias* results from a systematic difference in the way information on exposure or outcome is obtained from subjects. *Recall bias* refers to ill individuals recalling previous exposures differently from controls. A classic example of recall bias relates to studies of life events and depression, where individuals with depression tend to remember more negative life events, which non-depressed individuals forget or do not report. *Interview bias* involves bias in the interviewer's technique contingent on the interviewer knowing the disease status of subject; for example, if interviewers are aware of the case status of subjects, they may consciously or unconsciously put more or less effort into getting information on exposure variables. Control of bias may involve precautions taken in the study design, the choice of study population and the methods of data collection and can be addressed in the analysis of results.

Measurement in psychiatry

The easiest measuring instruments to use in psychiatry are self-report questionnaires, either as screening questionnaires for psychological distress, for example the General Health Questionnaire (GHQ) and the Malaise Scale, or for specific conditions, for example the Beck Depression Inventory (BDI), the Hospital Anxiety and Depression Scale (HADS) and the Center for Epidemiologic Studies Depression Scale (CES-D). The advantage of these scales is that they can be used as a continuous score for analysis, although they may not be normally distributed. The disadvantage is that they do not yield diagnoses.

Diagnostic agreement measured by kappa, intraclass correlations and Cronbach's alpha

It is important to ensure that such instruments are reliable. Reliability is of basically two types. One type of reliability refers to the internal consistency of the scale – whether the items on the scale are all measuring the same underlying dimension – assessed by high item–total score correlations, or *Cronbach's alpha*. The other type of reliability is retest reliability: does the test give the same score if it is repeated in the same individuals after a time interval? This is more relevant for dimensions that remain stable over time (e.g. neuroticism) rather than those that change (e.g. depressed mood).

Cronbach's alpha produces a value between 0 and 1. Typically, a value of 0.6 or higher is regarded as indicating adequate reliability, although ideally a value of 0.8 or higher should be sought. It is important to appreciate that the value of alpha will increase with the number of items in a scale.

Psychiatrists frequently measure various characteristics of individuals by the use of inventories, scales and measurements, making it necessary to measure the degree to which there is agreement between the raters undertaking the assessment. This is known as *inter-rater reliability*. The two most commonly used tests are described below.

The *kappa statistic* is used to assess agreement between two observers and an outcome. For example, if two psychiatrists both assess the same 30 patients for a diagnosis of schizophrenia, what is the extent of the agreement between them? Kappa assesses the reliability between the observers. Kappa can range from −1 to +1: −1 indicates no agreement at all between the two psychiatrists' diagnoses, 0 that there is agreement between the two psychiatrists but that this is not greater than the agreement we would expect by chance, and +1 that there is perfect agreement between the diagnoses for the psychiatrists. The significance of kappa is determined by a rule of thumb that states that agreement between observers should be at least 0.6 or higher if possible.

If there are more than two observers and with interval data, *intraclass correlation* (ICC) is used to assess agreement between observers. ICC compares the variability of different ratings of the same individual with the total variation in ratings across all the observers and individuals – that is, the variability of ratings for patient A are compared with the total variation across all the patients and all the observers. Like kappa, ICC produces an estimate that ranges from −1 to +1: it approaches +1 when there is little variability – that is, when the ratings for the individual patients are similar across observers. It is unusual to observe negative values in ICC.

In epidemiological research, case identification is important in counting the number of people with psychiatric disorders for assessing service need and changes in rates of illness. In most areas of psychiatry, there are no objective tests of psychiatric diagnosis such as raised blood sugar is for diabetes mellitus. Psychiatric diagnosis relies on clinical interview. In the past, different psychiatrists would have a range of diagnoses for the same patient. In other words, diagnoses had low reliability, where several assessors disagree in rating the same subjects; this is referred to as low inter-rater reliability. This has led to the development of

internationally agreed systems of diagnosis based on symptoms such as the *International Classification of Diseases*, version 10 (ICD-10), championed by WHO, and the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-IV), championed by the USA. Standardized interviews have been devised to gather data on symptoms in a systematic way in order to make diagnoses, such as the Present State Examination (PSE) and its successor Schedules for Clinical Assessment in Neuropsychiatry (SCAN), the Diagnostic Interview Schedule (DIS) and the Composite International Diagnostic Interview (CIDI). These interviews are often semi-structured, with screening questions that, if acknowledged, are followed up by additional probe questions. Such interviews require extensive training and are predicated on the interviewer being clinically trained. Increasingly, these interviews have been developed as fully structured, eliminating the need for interviewer discretion to enable their use by less expensive lay interviewers.

As well as reliability, an instrument needs to have validity – that is, does it measure accurately what it is supposed to measure? Face validity and content validity refer to whether the instrument appears to measure what you want to capture. Construct validity and discriminant validity relate to the instrument measuring the underlying construct but not other similar constructs. For instance, a valid measure of depression should not be measuring neuroticism or lack of social support. An important form of validity in psychiatry is criterion validity. In this, the sensitivity and specificity of a newly developed instrument are usually compared against a criterion, a so-called 'gold standard'. These gold-standard criteria are hard to find in psychiatry but often refer to a standardized psychiatric interview against which a simpler questionnaire is being assessed. Sensitivity of an instrument indicates the extent to which it identifies true positive cases of the illness. Those true cases of illness not identified by the test are false negatives and, together with true positives, comprise 100 per cent of cases. Specificity refers to the extent to which the instrument identifies true negatives, or non-cases of disease. Those who score positively on the screening instrument but who do not have the illness are false positives; this is often a fault of screening instruments for common mental disorders that tend to exaggerate the true prevalence of disorder. Together, true negatives and false positives comprise 100 per cent of non-cases. There are two other useful indices of the precision of an instrument. The positive predictive value (PPV) is the probability that a person actually has the illness given a positive test on the instrument. In contrast, the negative predictive value (NPV) is the probability that a person is free of the illness given an illness-free score on the instrument.

Increasingly, it is being recognized that psychiatric symptoms tend to fluctuate over time and may be transient indicators of more stable underlying latent traits or constructs. These underlying latent constructs may not be directly measurable and require sophisticated life-course analyses to identify them.[9] These underlying constructs may constitute endophenotypes that are linked more strongly to underlying genes than are symptoms or conventional psychiatric syndromes.

It is important to assess whether an association may be subject to bias. This may be due to *confounding*. A confounding factor is one that is associated with the exposure and also associated independently with the outcome and introduces a source of error that either diminishes or magnifies the true association. The association of the confounding factor with the outcome may not be causal, for example age and sex as confounding factors. It is important to adjust for confounding factors in analysis in order to see how much they may explain the association.

Descriptive and inferential statistics

Significance tests are used to evaluate numerical data, including categorical, ordinal and interval data, to test specific hypotheses between variables/measures. For example, imagine we are testing the hypothesis that, within a sample, females report higher rates of depression than males. Wherever there is a hypothesis, there is also a *null hypothesis* – which in this case would be that there are no differences between males and females in reports of depression.

Significance tests evaluate whether the results of a study are likely to have happened by chance. When epidemiological studies report the statistical significance of associations between variables, this is based on the *P* value. Levels of significance that are typically employed in epidemiological studies are 5 per cent ($P \leq 0.05$), 1 per cent ($P \leq 0.01$) and 0.1 per cent ($P \leq 0.001$), indicating, respectively, that the observed relationship between variables has a 5 in 100, 1 in 100 and 1 in 1000 probability of being due to chance. The lower the *P* value, the more certainty that the result is not due to chance. So, for example, if we found that females reported higher rates of depression than males and we found $P \leq 0.001$, we could be fairly certain that a gender difference did exist and we would accept our hypothesis that females report higher rates of depression than males. If the *P* value was greater than 0.05, this would indicate that there was no significant difference between males and females, and we would accept the null hypothesis. Note that the latter finding would be described as 'non-significant' in the text, not 'insignificant' – a frequently made lexical error.

In accepting and rejecting hypotheses, two types of error can be made. The first, known as a *type I error*, occurs when the null hypothesis is rejected even though it is true. In our example, this would mean that we would conclude that there was a gender difference in reports of depression, when actually there was not. The second, known as a *type II error*, occurs when we incorrectly reject the hypothesis and accept the null hypothesis. In our example, this would mean we would conclude that there was no gender difference, when there actually was. A type I error is most likely to occur

when the *P* value that we are willing to accept is set high; type I errors can be avoided by setting a lower *P* value and thereby reducing the probability of chance findings. Conversely, type II errors are more likely when we set a very low *P* value. The standard approach to balance the chance of making type I and type II errors is to set $P \leq 0.05$.

When a population has been sampled randomly, results can be generalized reasonably accurately to the entire population from which the sample was drawn by the use of confidence intervals (CIs). Confidence intervals define the range of values within which the true estimate for the parameter must lie. Confidence intervals are expressed as percentages, so the commonly used metric of 95 per cent confidence intervals indicates that we can be 95 per cent sure that the estimate lies within the intervals. Epidemiological research often reports the 95 per cent confidence intervals for odds ratios, derived from logistic regressions. For example, a reported logistic regression analysis of the 1958 British Birth Cohort found that females were more likely to have a depressive episode at 45 years of age than males, with the odds ratio reported as 1.85 and the 95 per cent confidence intervals as 1.33 to 2.57.[4] Although the overall odds for the statistical analysis is 1.85, the confidence intervals indicate that we have 95 per cent certainty that the true odds lies between 1.33 and 2.57 – that is, there is a 5 per cent margin of error. The significance of logistic regression analyses is evaluated by assessing whether the confidence intervals include 1. In the example above, being female has a significant effect on depressive episode at 45 years of age, as the confidence intervals range from 1.33 to 2.57. If the odds ratios crossed 1, for example from 0.97 to 2.57, then there would be no significant effect of being female on reports of depressive episode. If the confidence intervals covered a range below 1, for example from 0.26 to 0.94, then this would indicate support for the hypothesis that males were more likely than females to report a depressive episode.

The width of a confidence interval is affected by the size of the sample under investigation. If confidence intervals become very wide, this suggests that the estimates are unreliable, which is usually associated with a small sample that lacks power. Ideally, confidence intervals should be fairly narrow in their range, indicating that the sample is large enough for reliable estimates to be produced. Confidence intervals can be derived for many different types of estimate, including odds ratios, relative risk, proportions, means and percentages.

The main advantage of confidence intervals over *P* values, which makes them particularly attractive to epidemiologists, is that they permit the range of true estimates to be derived, enabling the effect size of the relationship between variables to be identified: confidence intervals enable the highest and lowest true estimate to be identified. In the example above, from the *P* value analyses, we would know only that we were more likely to see more females than males with depression; from the confidence intervals, we would know that we were likely to see between 1.33 and 2.57 times more females than males, which may alter how resources are allocated.

Specific statistical tests

Different statistical tests are used to determine relationships between variables of different types of data, such as categorical/discrete, continuous and ordinal. *Descriptive statistics* are used to describe the data collected in a study or experiment, as opposed to inferential statistics, which use the data to draw inferences about the population. Epidemiological studies typically employ a *within-subject* approach to analyses, focusing on describing and assessing differences across the subjects within the study population. For example, a birth cohort study may make comparisons between males and females, or between people in employment and people not in employment. All the subjects are from the same population. Another commonly used approach is the *between-subject* approach, which is a more experimental method in which two or more populations are compared with each other. For example, an RCT compares subjects who receive the intervention with subjects who do not receive the intervention.

Tests exist for both of these types of study, ranging from tests that simply provide estimates that describe the data (e.g. means, measures of central tendency) to tests that compare differences in outcomes between groups, such as males versus females. Many tests describe the associations between variables (e.g. that depression is associated with life events) but are limited in terms of being able to draw conclusions about the strength of association or causality between variables. This section describes some commonly used and encountered statistical tests within epidemiology.

Descriptive statistics for categorical data

The chi-squared (χ^2) measure of association/goodness of fit is based on evaluating the *distribution of observed data*. The test is most often used to analyse data from independent samples, as in the contingency table in Table 5.1; the term 'independent' means that each individual in the sample can appear in only one of the four cells. The test assesses *goodness of fit*, which is the test of the null hypothesis that there is no significant difference between the observed frequencies in the cells. In this example, the null hypothesis would be that there is no difference between current rates of

Table 5.1 Example contingency table showing the frequency of reports of current depression by previous depression

	No current depression	Current depression	Total (N)
No previous depression	50	40	90
Previous depression	30	60	90

depression for people who have and people who have not experienced depression before.

The most commonly used test is Pearson's χ^2 test, which is applied to normally distributed data. If the *P* value for the test is above the level of significance, which is conventionally set as $P \leq 0.05$, then the null hypothesis is accepted. If $P \leq 0.05$, then the null hypothesis would be rejected and we would conclude that there was a significant association between previous and current depression. One disadvantage of the χ^2 test for epidemiologists is that it does not tell us where the difference lies between the different cells in the contingency table: the test simply identifies that there is a significant association.

Descriptive statistics for continuous data

Most inferential statistical tests (see Inferential statistics, below) require that data be *normally distributed*. This means that the mean, mode and median have the exact same value and are located in the centre of a symmetrical distribution.

Variance measures the degree to which the value for each individual in the sample differs from the mean for the whole sample, thus assessing variability within the data. Variance is always a positive number, as any negative numbers, which can result from an individual scoring lower than the mean, are squared. Calculating variance is the first step in calculating the *standard deviation*, which is the square root of the variance. Standard deviations are usually reported wherever mean scores are reported, as they indicate the spread of the scores from the mean value.

Confidence intervals have two further popular uses in epidemiology. The first is as a means of describing the range within which the true value of a mean, proportion or median lies for the sample. Confidence intervals can also be derived for assessing the difference between two estimates of means, proportions and medians. This test assumes that the two samples have the same variance and are independent and that the data are normally distributed. The test involves deriving the estimate for each group, for example males and females, calculating the standard error of the difference between the two groups, and then deriving the confidence intervals by computing the estimate for the group of interest (e.g.) females $\pm$ 1.96 $\times$ standard error. 1.96 is the value to use for 95 per cent confidence intervals and 2.58 the value to use for 99 per cent confidence intervals.

Inferential statistics

Inferential tests are applied to the data to enable researchers to draw inferences and make assumptions about the population under study.

t-Tests are applied to test the difference in mean scores between two samples. These can be independent samples, such as the difference in the mean score on a depression scale for males and females, or they can be dependent, such as the difference in a depression scale score for the same sample of people before and after an intervention such as CBT. The *t*-statistic relies upon knowing the mean, standard deviation and number in each of the samples and requires that the data be normally distributed. The *P* value is used to evaluate whether the hypothesis that there is a significant difference between the mean scores is supported or refuted.

Where the data are not normally distributed and are from independent samples, the *Mann–Whitney U test* may be used instead. This test assumes only that the distributions from the two independent samples be of a similar shape, not that they be normally distributed. The test produces a U-statistic and a *P* value.

t-Tests can be used only where there are two samples to be compared. Where there are more than two samples, analyses of variance (ANOVAs) can be used. Unlike *t*-tests, which assess mean scores, an ANOVA compares the variance of scores across the different samples. ANOVAs are appropriate where data are interval, ratio or ordinal; independent; and normally distributed. ANOVAs produce an F-statistic and a *P* value.

Epidemiologists frequently employ *regression* techniques to develop complex models that examine the role of a predictor variable or several predictor variables on an outcome variable. For example, air pollution levels near schools may be used to predict asthma in children, and smoking behaviour may be examined in relation to mortality or heart disease. Such techniques are extremely useful for analysis of longitudinal datasets, where predictors from earlier in life can be examined as predictors of later outcomes.

Regression techniques exist for a range of types of outcome data, including continuous and categorical data, and data that measure the time taken for an event measured as a binary outcome to occur, for example mortality or disease onset. Continuous-outcome data, such as a score on a test, are analysed using *linear regression*; categorical data are analysed using *logistic regression*; and time-to-event binary data are analysed using *Poisson* or *Cox regression*, the latter also being known as *survival analysis*.

One of the main advantages of regression techniques is that confounding factors can be included in the model, along with the predictor variables. This means that the association between the predictor and the outcome can be examined after taking into account other important factors. For example, if a researcher examines the association between smoking behaviour and heart disease, they may want to adjust for other lifestyle factors such as diet and socioeconomic factors, as well as gender and age.

Another attraction for epidemiologists of regression techniques is that regression techniques result in estimates of the size of the effect of the predictor on the outcome. Knowing the size of the influence of the predictor on the outcome is important for public health.

Factor analysis is a data-reduction technique that examines the underlying structure of a set of variables, using the variance between the variables. Factor analysis identifies traits from the intervariable correlation matrix between variables. Factor analysis requires that data are continuous: categorical data cannot be factor analysed, and typically

this analysis is conducted on data that are collected using Likert scales.

Factor analysis is usually used to examine or identify unobserved latent traits or factors within datasets. For example, if the individual items on the Big 5 personality inventory were factor analysed, the result would be five different factors, or latent traits, of personality – neuroticism, conscientiousness, openness, extraversion and agreeableness. This is an example where the traits or factors are already known to the researcher. Factor analysis can also be used to identify factors where the researcher has not a priori identified the traits. For example, a number of continuous variables that measure attitudes towards antidepressant use could be factor analysed to identify the underlying structure of attitudes. It can be problematic to identify factors from the analysis; one approach, which has its critics and supporters, is to use *eigenvalues*, which identify the total variance that the factor accounts for in the analyses: factors with eigenvalues greater than 1 are typically thought to be worth examining. Factor scores can also be derived from the analysis, giving an estimate of each respondent's score for the factor, which can be used in further analysis.

Principal components analysis (PCA) is a similar data-reduction technique to factor analysis that also requires continuous data. PCA makes different mathematical assumptions about the nature of the relationship between the variables. PCA is usually used where the researcher wishes to reduce the number of variables in an analysis but to explain the same amount of variance, whereas factor analysis is used where underlying unobserved factors or latent constructs are of interest.

EPIDEMIOLOGY OF SPECIFIC PSYCHIATRIC DISORDERS

National surveys, such as the National Survey of Adults in Private Households carried out in the UK in 1993, 2000 and 2007, give useful estimates of the prevalence of psychiatric disorders (Table 5.2). In the USA, these include such surveys as the Epidemiologic Catchment Area (ECA) studies in the 1980s[10] and the National Comorbidity Survey (NCS) conducted in 1990–92 and its replication survey carried out in 2001–03 (Table 5.3). Rates of disorder differ between these studies, depending on the measuring instruments used, the study procedure, the choice of sample, and secular changes in rates of disorder. For instance, prevalence rates of psychiatric disorder were considerably higher in NCS than in

Table 5.2 Prevalence of psychiatric disorder in the UK general population: psychiatric morbidity among adults living in private households (2000)

	Total (%)	Women (%)	Men (%)
Neurotic disorder	16.4	19.4	13.5
Mixed anxiety and depression	8.8	10.8	6.8
Generalized anxiety disorder	4.4	4.6	4.3
Depressive episode	2.6	2.8	2.3
Psychotic disorder	0.5	0.5	0.6
Alcohol dependence	7.4	2.9	11.9

From Singleton *et al.*[13]

Table 5.3 Prevalence of psychiatric disorder in the National Comorbidity Survey Replication

Type of disorder	12-month prevalence*	Lifetime prevalence**
Panic disorder	2.7	4.7
Agoraphobia without panic	0.8	1.4
Specific phobia	8.7	12.5
Social phobia	6.8	12.1
Generalized anxiety disorder	3.1	5.7
Post-traumatic stress disorder	3.5	6.8
Obsessive–compulsive disorder	1.0	1.6
Major depressive disorder	6.7	16.6
Dysthymia	1.5	2.5
Bipolar I and II	1.5	3.9
Alcohol abuse	3.1	13.2
Alcohol dependence	1.3	5.4
Drug abuse	1.4	7.9
Drug dependence	0.4	3.0

*Kessler *et al.*;[14] **Kessler *et al.*[15]

the earlier ECA studies – it is unlikely that these changes represented a real increase in prevalence. More recent prevalence surveys include the WHO World Mental Health Survey using the CIDI across 28 countries and finding a broad range of prevalence estimates for mood disorders; the WHO Collaborative Study of Psychological Problems in Primary Care;[11] and the European Study of the Epidemiology of Mental Disorders project (ESEMED).[12]

A brief summary of epidemiological data for specific diagnoses follows.

Schizophrenia

The lifetime prevalence is 4.3 per cent, lifetime risk is 0.72 per cent and adult point prevalence is 4.5 per cent. Incidence of schizophrenia is 15.2 in 100 000. The male/female ratio is 1.4 : 1. Compared with native-born individuals, migrants have an increased incidence and prevalence of schizophrenia. Exposures related to urbanicity, less advantaged economic status and certain latitudes (increased in winter births) are associated with increased risk. People with schizophrenia have a two- to threefold increased risk of dying prematurely compared with the general population (median standardized mortality ratio = 2.6 for all-cause mortality).[16]

Manic–depressive psychosis (bipolar disorder)

Lifetime prevalence of bipolar I disorder in the NCS replication study is 1.0 per cent; 6-month prevalence is 0.6 per cent. Lifetime prevalence of bipolar II disorder is 1.1 per cent and 12-month prevalence is 0.8 per cent.[17] Onset is in late teens to late thirties. Regarding the co-morbidity of lifetime bipolar disorder, 63–87 per cent of patients have a lifetime anxiety disorder, and 35–60 per cent a lifetime substance use disorder. It is related inversely to age, educational level and previous marriage. It is unrelated to sex, ethnicity and income.

Major depression

Lifetime prevalence in the NCS replication study is 16.2 per cent (95% CI: 15.1% to 17.3%). The 6-month prevalence is 6.6 per cent (95% CI: 5.9% cent to 7.3%).[18] The male/female ratio is 2 : 1. Major depressive disorder (MDD) is highly co-morbid, with a 59 per cent risk of a lifetime anxiety disorder and 24 per cent risk of a substance use disorder. The median age of onset is 32 years (interquartile range 19–44 years). Risk factors include negative (loss) life events, low socioeconomic status, physical illness (e.g. coronary heart disease, cancer) and childhood adversities.

Suicide

The prevalence is 16.7/100 000 annually. Suicide makes up 1.5 per cent of all deaths and is the fourteenth leading cause of death.[19] The lifetime prevalence of suicide attempts is 0.4–5.1 per cent. The male/female ratio is between 3 : 1 and 7.5 : 1. There is an increase in non-fatal self-harm in people aged 12–20 years, peaking at age 16 years. Risk factors for completed suicide include male, adolescent, older adult, certain ethnic groups, unemployment, certain psychiatric disorders (depression, impulse-control disorders, alcohol/substance use disorders, psychotic disorders, personality disorder), hopelessness, stressful life events and certain occupations (e.g. military personnel, doctors, pharmacists).

Generalized anxiety disorder

The lifetime prevalence is 4.3, 5.9 per cent. The 12-month prevalence is 1.2–1.9 per cent. The male/female ratio is 1 : 2. Onset is usually before age 25 years. Some 66 per cent of patients have co-morbid major depression; a similar proportion have other anxiety disorders.[20]

Panic disorder

The adult prevalence of panic disorder with agoraphobia is 2–3 per cent; the prevalence is 1 per cent without agoraphobia. The lifetime prevalence is 1–2 per cent. The male/female ratio is 1 : 2. Onset is in late adolescence or early adulthood. Fifty per cent of patients develop depression during their lifetime, and there is an increased risk of suicide. Risk factors include heritability (40%), anxious temperament, anxiety sensitivity, neuroticism, early life trauma and recent life events.

Specific phobias

The adult prevalence is 2–5 per cent. The lifetime prevalence is 7–11 per cent. The male/female ratio is 1 : 2. Onset is usually at 5–8 years.

Agoraphobia

The adult prevalence is 2–3 per cent. The lifetime prevalence is 1.6 per cent. The male/female ratio is 1 : 4. Onset is usually around 18–35 years. It is associated with co-morbid major depression.

Social phobia

The adult prevalence is 4–6 per cent. The lifetime prevalence is 13 per cent. The male/female ratio is 1 : 1.5. Onset is usually in the teens (from 11 years) to early twenties (80% by 20 years). Risk factors include heritable components related to behavioural inhibition and neuroticism.[21]

Obsessive–compulsive disorder

The lifetime prevalence is 2.5 per cent. The 12-month prevalence is 0.7–2 per cent. The male/female ratio is between 1 : 1.5 and 1 : 1. There is increasing incidence with

age from 15 years to 35 years; the mean age of onset is 19.5 years. The most common co-morbid conditions are anxiety disorders (75.8%), mood disorders (63.3%), impulse-control disorders (55.9%) and substance use disorders (38%). Sixty-five per cent of patients report severe role impairment.[22]

Post-traumatic stress disorder

The lifetime prevalence is 8 per cent. The 12-month prevalence is 1 per cent. The male/female ratio is 1 : 2.

Alcohol use disorders

The lifetime prevalence is 13.8, 23.5 per cent. The adult 12-month prevalence is 6.8, 7-7 per cent. The male/female ratio is 2 : 1. Alcohol use onset is usually in adolescence and peaks in late adolescence or early adulthood. Risk factors include alcohol-relieving negative affect, stimulating brain reward centres, pharmacological vulnerability, deviance proneness, heritability, parental substance abuse and peer influences. There is co-morbidity with all anxiety and mood disorders, particularly mania, antisocial behaviour disorders and drug use.

Drug dependence

The adult 12-month prevalence is 6.1, 7.5 per cent. The lifetime prevalence is 15 per cent. Eighteen per cent of people with a lifetime major depressive disorder have a substance use disorder. There is a higher risk for bipolar disorder.

Anorexia nervosa

The incidence rate is between 4.7 and 8.3 per 100 000. The lifetime prevalence is 1.2–2.2 per cent. The 12-month prevalence is 0.37 per cent. The 12-month prevalence of bulimia nervosa is 1.5 per cent. The male/female ratio is 1 : 8.[23]

Alzheimer's disease

The incidence rate is 1 per 1000 person-years in people aged 60–64 years, and 25 per 100 person-years in people over 85 years of age. The age-standardized prevalence rate is 4.4 per cent in people over 65 years and 22.2 per cent in people over 90 years. It is more frequent in people of advanced age, females, carriers of the apolipoprotein E4 gene, current smokers, people with a low education, income, and lower occupational status, and family history.

KEY POINTS

- Epidemiological methods can be used to describe the burden of illness in a population, to contribute to disease aetiology, to test the effectiveness of health interventions and to assess need for services.
- Epidemiological studies include cross-sectional (prevalence and case–control) studies and longitudinal (incidence and cohort) studies. Longitudinal studies are better than cross-sectional studies for assessing aetiology.
- Studies may be subject to bias. Observation bias occurs where there is a systematic difference in the way information on exposure or outcome is obtained from subjects. Selection bias relates to non-random selection of subjects. Recall bias refers to ill individuals recalling previous exposures differently from controls.
- The reliability of an instrument refers to the agreement between items making up a scale (internal consistency) and the stability of results over time in the same individuals, assuming no change in the construct measured (retest reliability).
- The criterion validity of an instrument measures how well it compares with an existing gold standard instrument. The sensitivity of an instrument is an index of how well it identifies true positive cases. The specificity of an instrument measures how well it identifies true negatives – that is, individuals without disease. The positive predictive value is the probability that a person actually has the illness given a positive test on the instrument. The negative predictive value is the probability that a person is free of the illness given an illness-free score on the instrument.
- A confounding factor is one that is associated with the exposure and also associated independently with the outcome and introduces a source of error, which either diminishes or magnifies the true association. Confounding can be dealt with either in the design or in the analysis of a study.
- A type I error occurs when the null hypothesis is rejected, even though it is true. A type II error occurs when we incorrectly reject the hypothesis and accept the null hypothesis.

REFERENCES

1. Eaton WW (1999) Psychiatric epidemiology. In H Freeman (ed.) *A Century of Psychiatry*. London: Mosby.
2. Kety SS, Rosenthal D, Wender PH, Schulsinger F, Jacobsen B (1975) Mental illness in the biological and adoptive families of adopted individuals who have become schizophrenic: a preliminary report based on psychiatric interviews. In RR Fieve, D Rosenthal, H Brill (eds) *Genetic Research in Baltimore Psychiatry*. Baltimore, MD: Johns Hopkins Press.
3. Wadsworth M, Kuh D, Richards M, Hardy R (2006) Cohort profile: the 1946 National Birth Cohort (MRC National Survey of Health and Development). *International Journal of Epidemiology* **35**: 49–54.
4. Clark C, Rodgers B, Caldwell T, Power C, Stansfeld S (2007) Childhood and adulthood psychological ill health as predictors of midlife affective and anxiety disorders: the 1958 British Birth Cohort. *Archives of General Psychiatry* **64**: 668–78.
5. Marmot MG, Smith GD, Stansfeld S, *et al.* (1991) Health inequalities among British civil servants: the Whitehall II study. *Lancet* **337**: 1387–93.

6. Van den Berg B, Grievink L, *et al*. (2008) Risk factors for physical symptoms after a disaster: a longitudinal study. *Psychological Medicine* **38**: 499–510.
7. Hennekens CH, Buring JE (1987) *Epidemiology in Medicine*. Philadelphia, PA: Lippincott Williams & Wilkins.
8. Lee W, Bindman J, Ford T, *et al*. (2007) Bias in psychiatric case–control studies: literature survey. *British Journal of Psychiatry* **190**: 204–9.
9. Colman I, Ploubidis GB, Wadsworth ME, Jones PB, Croudace TJ (2007) A longitudinal typology of symptoms of depression and anxiety over the life course. *Biological Psychiatry* **62**: 1265–71.
10. Regier DA, Boyd JH, Burke JD, Jr, *et al*. (1988) One-month prevalence of mental disorders in the United States: based on five Epidemiologic Catchment Area sites. *Archives of General Psychiatry* **45**: 977–86.
11. Ustun TB, Sartorius N (1995) *Mental Illness in General Health Care*. Chichester: John Wiley & Sons.
12. Alonso J, Angermeyer MC, Bernert S, *et al*. (2004) Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatrica Scandinavica* **109**: 21–7.
13. Singleton N, Bumpstead R, O'Brien M, Lee A, Meltzer H (2001) *Psychiatric Morbidity Among Adults Living in Private Households 2000*. London: Office for National Statistics.
14. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE (2005) Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry* **62**: 617–27.
15. Kessler RC, Berglund P, Demler O, *et al*. (2005) Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry* **62**: 593–602.
16. McGrath J, Saha S, Chant D, Welham J (2008) Schizophrenia: a concise overview of incidence, prevalence, and mortality. *Epidemiology Review* **30**: 67–76.
17. Merikangas KR, Akiskal HS, Angst J, *et al*. (2007) Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. *Archives of General Psychiatry* **64**: 543–52.
18. Kessler RC, Berglund P, Demler O, *et al*. (2003) National Comorbidity Survey Replication: the epidemiology of major depressive disorder – results from the National Comorbidity Survey Replication (NCS-R). *Journal of the American Medical Association* **289**: 3095–105.
19. Nock MK, Borges G, Bromet EJ, *et al*. (2008) Suicide and suicidal behavior. *Epidemiology Review* **30**: 133–54.
20. Tyrer P, Baldwin D (2006) Generalised anxiety disorder. *Lancet* **368**: 2156–66.
21. Stein MB, Stein DJ (2008) Social anxiety disorder. *Lancet* **371**: 1115–25.
22. Ruscio AM, Stein DJ, Chiu WT, Kessler RC (2008) The epidemiology of obsessive–compulsive disorder in the National Comorbidity Survey Replication. *Molecular Psychiatry* [forthcoming].
23. Hoek HW (2006) Incidence, prevalence and mortality of anorexia nervosa and other eating disorders. *Current Opinions in Psychiatry* **19**: 389–94.

Chapter 6

How to practise evidence-based medicine

Stuart Carney

INTRODUCTION

The practice of evidence-based medicine (EBM), which is the integration of the best available evidence of clinical experience and patient values in patient or clinical decision-making, can be broken down into five steps:[1]

1 Translation of uncertainty to an answerable question.
2 Systematic retrieval of the best available evidence.
3 Critical appraisal of the evidence.
4 Application of the results in practice.
5 Evaluation of performance.

Put another way, to practise EBM, you need to do the following:

1 Ask structured clinical questions.
2 Access the best available evidence.
3 Appraise the evidence.
4 Apply the evidence.
5 Assess performance and patient outcomes.

This chapter systematically considers each of these five steps. It builds upon the earlier chapters in this part on epidemiology and research methods and statistics.

STEP 1: ASK QUESTIONS

Many questions emerge in the course of clinical practice. One model for translating uncertainty into an answerable question is to break it down into four parts known by the acronym PICO:

- Patient
- Intervention (or E for exposure)
- Comparison
- Outcome.

For example, imagine you are faced with the following clinical situation:

Mrs B is a 31-year-old single mother with three primary school-aged children. She has suffered from recurrent episodes of depression since the age of 23 years. The latest depressive episode is severe, she is not eating and she is barely drinking. The treatment options include electroconvulsive therapy (ECT) and antidepressant medication.

We might ask: is ECT more effective and safer than antidepressants in the treatment of a woman with a severe depressive episode? Using the PICO model, we could break down this question into four parts:

- *Patient:* woman with a severe depressive episode
- *Intervention:* ECT
- *Comparison:* antidepressant treatment
- *Outcome:* remission of the depressive symptoms, side effects.

STEP 2: ACCESS THE EVIDENCE

Types of clinical question

Therapy questions, such as ECT versus antidepressants for the treatment of severe depression, are particularly common in clinical practice. However, clinical practice throws up other types of question: diagnosis, prognosis, aetiology and harm. Epidemiology, the study of the distribution and determinants of disease, provides a framework for addressing each type of question. Different types of study are best suited to answering specific types of question. It is essential to determine what type of question is being considered before searching the literature. The more common types of clinical question are described below.

Diagnostic questions (distribution of disease)

The medical paradigm is predicated on making a diagnosis. We seek to determine whether a patient is suffering from an illness and, if so, the exact nature of that illness. This decision-making process is informed by knowledge of the distribution of possible illnesses in the target population. Diagnostic instruments and tests are often used to assist in this process as the clinician seeks to confirm or refute

hypotheses (differential diagnoses). As clinicians, we need to stay abreast of the evidence relating to prevalence and the accuracy of diagnostic tests. Cross-sectional studies are used to address these questions.

Prognosis questions

When patients ask what will happen to them, they are asking about their prognosis. Prognosis describes the likely course in a person with specific characteristics.

Aetiology, harm and therapy questions (determinants of disease)

What caused this disease? Is this drug an effective treatment for this disease? These are examples of aetiology, harm or therapy questions. Aetiology questions seek to ascertain what factor or factors caused a diagnosed disease, while therapy and harm questions focus on interventions that may change the course of the disease.

Hierarchy of evidence

The optimal study design to address a particular question depends on the type of question being asked. The pitfalls outlined in Chapter 4 can threaten the validity and importance of any clinical study, but certain designs are potentially better than others. The hierarchy of study architecture for each of the three groups of clinical questions is shown in Table 6.1.

It is also worth reflecting on two specific types of study, which are particularly important to the practice of EBM – systematic reviews and *N*-of-1 studies.

Systematic reviews

In 1979, Archie Cochrane, a British epidemiologist, lamented the failure of the medical profession to 'organise a critical summary, by specialty or subspecialty, adapted periodically, of all relevant randomised controlled trials'.[2] Thirteen years later, the Cochrane Collaboration was founded to 'Prepare, maintain and promote the accessibility of systematic reviews of the effects of health care interventions'. Although systematic review methodology is most advanced for the synthesis of randomized controlled trials (RCTs), the same principles can be applied to any study design. Consistent results from valid and important studies synthesized in a systematic review can provide the best available evidence for all clinical questions.

N-of-1 studies

Evidence-based medicine focuses on the care of the individual patient, while evidence-based healthcare relates to the effects of exposures in groups. It may not be appropriate or possible to apply population-based observations to the care of a particular patient. In certain circumstances, it may be possible to undertake research in partnership with an individual consenting patient. For conditions that are chronic

Table 6.1 Hierarchy of study designs

	Determinants of disease	Prognosis questions	Distribution of disease	
	Therapy, aetiology or harm questions		Diagnosis questions	
			Surveys (prevalence)	Evaluation of diagnostic instruments
1	*N*-of-1 study	Systematic review of cohort studies	Systematic review of cross-sectional surveys	Systematic review of cross-sectional diagnostic studies
2	Systematic review of RCTs	Individual cohort study	Single survey	Individual cross-sectional diagnostic study
3	Individual RCT	Follow-up of untreated control patients in an RCT	Expert opinion	Expert opinion
4	Controlled clinical trial	Case series		
5	Cohort study	Expert opinion		
6	Case–control study			
7	(Cross-sectional study)			
8	(Ecological survey)			
9	Case series			
10	Expert opinion			

RCT, randomized controlled trial.

and the effects of an intervention are transient, we could conduct an *N*-of-1 study.

By definition, the observations will be applicable to the patient under consideration. In an *N*-of-1 study, the patient is sequentially randomized to intervention and placebo and the outcomes assessed. Such a prospective design reduces the risk of performance and measurement bias and allows an assessment of the relative efficacy of an intervention. It is argued that the *N*-of-1 study design is primary for questions of therapeutic effect (therapy).[3]

Searching for the evidence

Having formulated a structured clinical question and determined the optimal study design, it should now be possible to search efficiently for the best available evidence. Developments in information technology, notably the development of electronic databases, have made the practice of EBM possible.

Table 6.2 lists some of the more common databases that may be of use when practising EBM. There are an increasing number of meta-resources, including clinical practice guidelines and evidence-based clinical decision support tools. These provide pre-appraised material addressing common clinical questions. To answer questions not covered by these resources, we must appraise systematic reviews or the original literature.

Electronic databases such as PubMed and Embase are useful resources for searching the scientific literature. PubMed is provided free by the US National Library of Medicine and is North American in focus, while Embase catalogues more European journals. The Cochrane Library includes the Cochrane Reviews, which are regarded as the gold standard for systematic reviews, and the Cochrane Controlled Trials Register (CCTR). The CCTR contains randomized and controlled trials drawn from sensitive searches of a number of databases, including Embase and PubMed.

Most electronic databases use Boolean logic to combine search terms. Table 6.3 provides some useful tips for searching. Returning to the clinical question about ECT, if we wish to find as many of the therapy studies as possible, then we could combine terms using the Boolean operator OR. As ECT is also known as electroshock therapy and electroconvulsive therapy, it would be sensible to search for these terms as well. A comprehensive search for papers on ECT could include the following search terms:

'electroconvuls*' or 'electro-convuls*' or 'electroshock*' or 'electro-shock*' or 'convuls*' or 'ECT'

As we are interested only in ECT for the treatment of a patient with depression, we could combine the ECT terms with search terms for depression. 'Depressive episode' is a synonym for 'depression', but both terms could be identified by truncating depression to 'depress*'. Similarly we could search for synonyms of 'antidepressants'. As we wish to restrict the search to these three criteria, we could use the Boolean operator AND.

Table 6.2 Databases useful in the practice of evidence-based medicine

Critically appraised databases/resources	Other electronic databases
Clinical Practice Guidelines	Cochrane Library
Clinical Evidence	PubMed
Evidence-based Medicine	Embase
Evidence-based Mental Health	PsycINFO
	Biological Abstracts
	Cinahl

Table 6.3 Electronic database search terms and tips

Search term	Notes
AND	Includes only those studies that contain both terms of interest, e.g. both 'depression' and 'ECT'
*	Can be used to search for all studies containing words beginning with the letters before the asterisk (*), e.g. 'depress*' will identify all studies with the words 'depression', 'depressive', etc. anywhere in the citation or abstract
OR	Includes all studies that contain term a, or term b, or both a and b
?	Can be used when various spellings are used for a term, e.g. '?etiology' will search for both 'aetiology' and 'etiology'
NOT	Includes all studies that contain term a, except for those that also contain term b
"..."	Can be used when a specific piece of text is to be searched for, e.g. 'University of Oxford' as a search term will identify only papers with this exact text and not those referring to 'Oxford University'

STEP 3: APPRAISE THE EVIDENCE

Critical appraisal addresses three questions: is the research applicable to my patient, is it valid, and is it important? Before considering the pitfalls in the design and execution of studies, it is worth pausing to consider how to efficiently extract the necessary data from papers.

Read a paper in less than 15 minutes

As busy clinicians, we often have to make rapid judgements. We do not have time to wade through piles and piles of research papers. With practice, it becomes easier to identify quickly which research papers are likely to address our clinical questions.

The title of the paper often gives clues as to whether it is likely to address the question of interest. More detail is given in the abstract. The title and the abstract help us to assess whether the paper is applicable to our patient – that is, whether it addresses our focused clinical question and uses an appropriate study design. Having spent no more than a minute looking at the paper, we should be able to make a judgement as to whether the paper is worth more detailed consideration.

The introduction rarely provides any useful information for critical appraisal, although the hypothesis under consideration or the aim of the study is usually described in the final paragraph. The methods and results sections should provide us with the data needed to critically appraise the paper. Having critically appraised the paper, there should be no need to read the discussion section, as we can draw our own conclusions. This whole process should take no more 15 min.

The GATE frame: critical appraisal with pictures[4]

Rod Jackson and colleagues have proposed a simplified approach to critical appraisal.[4] Inspired by the symbols on the Sony PlayStation®, they have created a frame (Graphic Appraisal Tool for Epidemiological studies, or GATE) comprising a triangle, circle, square and cross, upon which the key components of epidemiological studies can be hung (Figure 6.1).

The triangle represents the three levels of selecting participants for a study. For the ECT study, the source population would comprise patients with severe depression, the eligible population, those who satisfy the inclusion and exclusion criteria, and the sample those who agreed to take part. The circle is divided into two segments: those receiving the intervention (or exposure) and those receiving the comparison. The square can be broken down into a two-by-two table and is useful when we are interested in dichotomous outcomes (remission or no remission).

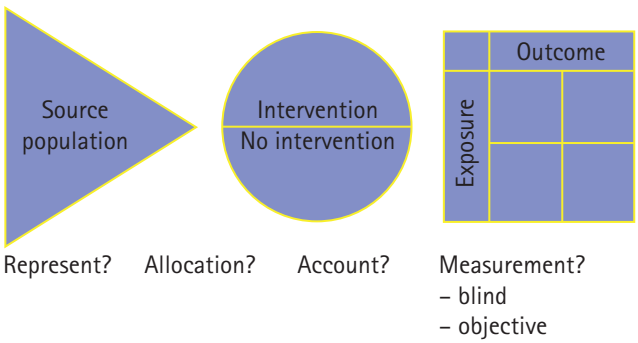

Figure 6.1 The GATE frame

The triangle, circle and square serve as a template to guide the extraction of key data about where the participants came from and what happens to them during the course of the study. In parallel, we need to consider the potential pitfalls in the design and execution of the study. The GATE frame proposes that these can be summarized by the acronym RAAMbo.

- *Represent:* this asks the question 'Is my patient sufficiently similar to the participants in this study?' This is also known as generalizability, external validity or applicability.
- *Allocation:* for therapy studies, this considers how the participants were allocated to the intervention and comparison groups. Randomization with allocation concealment seeks to reduce the risk of selection bias, by achieving balance between the two groups with respect to known and unknown risk factors. Imbalance between the intervention and comparison group results in confounding. In prognostic studies, the cohorts are defined by whether or not they have a particular characteristic or exposure. Strategies may be employed to control for known prognostic factors when comparing the two groups in both the assembly (e.g. matching) of the groups and the analysis.
- *Accounted:* ideally, by the end of the study all participants should be accounted for, but this is rarely the case and typically some participants will drop out, switch treatment or be lost to follow-up. To maintain the integrity of the original allocation, attempts should be made to analyse all participants in the group to which they were allocated. This may require us to make judgements about how to handle dropouts (e.g. last observation carried forwards, best case/worst case scenario). Analysing all participants in the group to which they were originally allocated is known as 'intention-to-treat' analysis. Studies that experience a significant difference in dropouts across the two or more arms are prone to bias.
- *Measurement:* measuring outcomes typically requires some subjective judgement, and there is a risk that the outcome assessors may apply different thresholds, depending on whether the participant was allocated to a particular intervention. Blinding is a strategy to reduce the risk of measurement bias, and it is important that the study authors describe who was blinded and how blinding was achieved. There is less scope to make judgements if the outcome measures are objective (e.g. death).

The GATE frame provides a model for extracting data and making judgements about the applicability and validity of the study. If we are satisfied that any failings in the design or execution of the study are not catastrophic and that the study provides the best available evidence, we can

turn our attention to the numbers or importance of the study. When dichotomous outcome measures are considered (event or no event), it is possible to calculate measures of effect such as absolute risk reduction (ARR), risk ratio (RR), relative risk reduction (RRR) and number needed to treat (NNT). These terms are described in Table 6.4. The terms ARR and RRR are used when the intervention reduces the risk of adverse events. Absolute benefit increase (ABI) and relative benefit increase (RBI) are more appropriate when considering positive outcomes.

Is it a poor-quality study or a poorly reported study?

There is a danger with critical appraisal that we become cynical about research. Most enterprises involving large numbers of people rarely go exactly to plan. Even the most perfectly designed study will encounter problems in its execution. It is common for participants to withdraw and no longer be available for follow-up.

Mindful that word limits and copy space present researchers and editors with challenges, even the most fundamental information about the conduct of the studies is often omitted. Therefore it is difficult to judge whether the study is of poor quality or just poorly reported.

In an attempt to ensure that readers of the scientific literature are equipped with all of the necessary information, editors and researchers have agreed standards for reporting research. The Consolidated Standards of Reporting Trials (CONSORT) statement recommends that authors and journal editors follow a 22-item checklist when reporting RCTs and use a flow diagram to describe the progress of study participants.[5] Although only 20 per cent of high-impact medical journals refer to the statement in their advice to authors, the quality of reporting of RCTs is improved in journals that have adopted it.[6,7]

Like the CONSORT statement, the Quality of Reporting of Meta-Analyses (QUORUM) statement describes the critical steps in a systematic review and meta-analysis of RCTs.[8] Guidance to authors and editors is not confined to RCTs and meta-analyses: there are also Standards for Reporting of Diagnostic Accuracy (STARD).[9]

STEP 4: APPLY THE EVIDENCE

The best available evidence is only one factor that should be taken into account for patient or clinical decision-making. There are clearly clinical considerations such as the severity of the illness and co-morbid risk factors. Patients, like clinicians, draw upon experience and have opinions and values. We make decisions in the context of healthcare policy, such as what is available in our particular health economy. There may be specific legal considerations that also need to be taken into account, such as the use of ECT when a patient refuses treatment or lacks capacity to give consent.

Clinical expertise is the ability to integrate research evidence and patients' circumstances and preferences in order to help patients arrive at optimal decisions.[1]. This is otherwise known as the X factor (Figure 6.2).[4]

Table 6.4 Effect measures

Term	Definition	How is it calculated?
Control event risk (CER)	Proportion with the event of interest in the control group	Number of events (controls)/total number in control group
Experimental event risk (EER)	Proportion with the event of interest in the intervention group	Number of events (intervention)/total number in intervention group
Absolute risk reduction (ARR)	Difference in risk between the control and experimental groups	ARR = CER – EER
Number needed to treat (NNT)	Number of patients that have to be treated with the intervention in order to prevent one additional person experiencing the adverse event; conventionally, NNT is rounded up	NNT = 1/ARR
Risk ratio (RR)	Risk of experiencing the event in the experimental group relative to the control group	RR = EER/CER
Relative risk reduction (RRR)	Proportion of events that would have been avoided in the control group had they been allocated the intervention	RRR = 1 – RR = ARR/CER

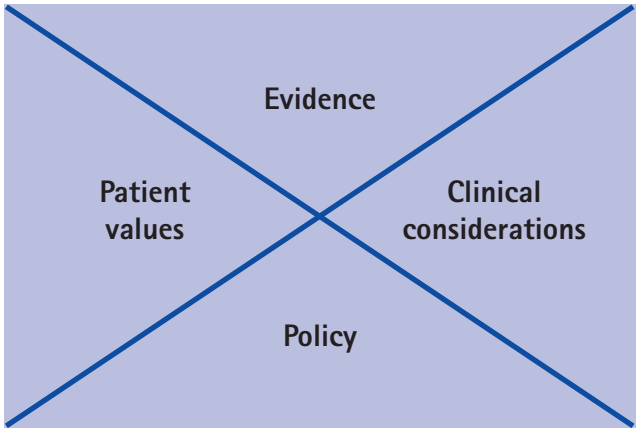

Figure 6.2 eXpertise

STEP 5: ASSESS PERFORMANCE AND PATIENT OUTCOMES

The most neglected step in the practice of EBM is the assessment of performance. As we strive to improve patient safety and care, we must assess whether patients have been helped and then consider what lessons can be learned from the care of individual patients. It may be the case that our review of the evidence focusing on the care of an individual patient highlights opportunities to improve the care of other patients.

REFERENCES

1. Straus S, Richardson W, Glasziou P, *et al.* (2005) *Evidence-Based Medicine: How to Practise and Teach*, 3rd edn. Edinburgh: Churchill Livingston.
2. Cochrane AL (1979) 1931–1971: a critical review, with particular reference to the medical profession. In Teeling-Smith G, Wells N, Office of Health Economics (eds) *Medicines for the Year 2000*. London: Office of Health Economics.
3. Guyatt GH, Haynes RB, Jaeschke RZ, *et al.* (2000) Users' guides to the medical literature XXV: evidence-based medicine – principles for applying the users' guide to patient care. *Journal of the American Medical Association* **284**: 1290–96.
4. Jackson R, Ameratunga S, Broad J, *et al.* (2006) The GATE frame: critical appraisal with pictures. *Evidence-Based Medicine* **11**: 35–8.
5. Moher D, Schulz KF, Altman DG, for the CONSORT Group (2001) The CONSORT statement: revisited recommendations for improving the quality of reports of parallel-group randomized trials. *Lancet* **357**: 1191–4.
6. Altman DG (2005) Endorsement of the CONSORT statement by high impact medical journals: survey of instructions for authors. *British Medical Journal* **330**: 1056–7.
7. Plint AC, Moher D, Morrison A, *et al.* (2006) Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Medical Journal of Australia* **185**: 263–7.
8. Moher D, Cook DJ, Eastwood S, *et al.* (1999) Improving the quality of reports of meta-analyses of randomized controlled trials: the Quorum statement. *Lancet* **354**: 1896–900.
9. Bossuyt PM, Reitsma JB, Bruns DE, *et al.* (2003) Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *British Medical Journal* **326**:41–4.
10. Guyatt G, Cook D, Haynes B (2004) Evidence based medicine has come a long way. *British Medical Journal* **329**: 990–91.

Chapter 7

Psychological assessment and psychometrics

Ann Mortimer

INTRODUCTION

The assessment of mental functions such as memory, concentration, language, intelligence, reasoning and judgement is a cornerstone of psychiatric practice. Impairment is germane to important disorders such as learning disability and the dementias, while there is some relevance to just about every other major diagnosis in psychiatry. It has been demonstrated repeatedly that poor cognitive ability is associated with both increased risk for mental disorder, and poorer personal and social function. A thorough understanding of cognitive function can, therefore, illuminate patient formulation to a surprising degree and may inform management in terms of how the patient can be expected to function in the future. This is particularly consequential for rehabilitation.

This chapter comprises an exposition of cognitive assessment beyond the usual few questions traditionally taught as part of mental state examination and the Mini-Mental State Examination (MMSE).[1] Comprehensive cognitive assessment starts with the taking of a history and examination of the rest of the mental state before formal screening for impairment. This in turn directs attention to more specific areas, or further psychometric tests may be indicated. The recognition of patterns of impairment informs diagnostic possibilities: common associations between specific profiles and the diagnoses which they imply will be given.

TAKING A HISTORY

Intelligence is a reasonable synonym for overall cognitive ability. Even without standard formal measures of intelligent quotient (IQ), a rough estimate can be made from the patient's history and general presentation. It should be possible to distinguish patients likely to be learning disabled (IQ 70, or IQ 75 where there is functional disability) with reasonable confidence.

IQ can be compromised by any significant insult to the brain, from gestation onwards. A history of adverse events in utero, birth complications, head injury, infections and systemic disorder in the preschool years or beyond may all be relevant here. The timescale of achievement of developmental milestones is useful information if available.

The patient's family background is also significant. Children brought up within very socially disadvantaged families or subjected to neglectful or abusive parenting are unlikely to learn as they should, through lack of appropriate stimulation and encouragement. There was previously a diagnosis, now obsolete, of 'subcultural subnormality', in which children assumed to be of ordinary intelligence were unable to manifest the intelligence because their parents were either very unintelligent or neglectful, or both.[2–4] These children could not learn from their parents, and they grew up with poor interest in and expectations of their achievement from family, schools and themselves. Presently, growing up with social and cultural deprivation potentially leads to a failure to access and benefit from learning opportunities, and a lack of interest in achieving potential in adulthood. Relevant enquiries about family poverty and disadvantage include the patient remembering being cold, hungry or in the dark, disconnection of utilities, the absence of a parent, separation from the parents, parenting from non-parents, family criminality, periods in foster care or on the Child Protection Register, and any kind of physical or emotional neglect and abuse.

These matters may impact equally, or more, on personality development in addition to compromising intellectual ability and potential. Although rather beyond the remit of this chapter, it is worth pointing out that deviance in cognition (as opposed to impairment) is one of the areas in which DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders*, 4th edn) personality disorder manifests itself, cognition here defined qualitatively as ways of perceiving and thinking about oneself, other people and events.

School performance can be an important indicator of intelligence. Was the patient educated in remedial classes or in an educational facility for children with learning problems? Did the patient have a Statement of Special Educational Needs from the local education authority? If so,

why, and what provision was made (e.g. one-to-one support)? Did the patient truant, or leave school early or without qualifications? What was the patient's attitude to leaning? Did the patient have a favourite subject? Again, this information has implications for a number of other issues apart from cognition, such as personality, cognitive style, attitudes and values.

It is unusual for patients with learning disability to be able to read, comprehend and write competently, and to be numerate. Establish everyday numeracy: the patient may be asked whether they can ensure that they pay the right amount for purchases and receive the correct change afterwards. Do they need help with filling in forms? Do they read regularly? If so, what can they read and understand (e.g. the tabloid press)?

Higher or further education, successfully completed, comprises additional evidence of cognitive ability. Even the patient who failed to complete secondary education may demonstrate the ability to learn new skills, such as driving a vehicle. Cognitive assessment naturally encompasses ability as well as deficit: what patients like to do and what they are good at. Such attributes are useful in addressing rehabilitative aspects of mental disorder. Interests, sports and hobbies may constitute evidence of intellectual aptitude or motor abilities, or both.

Apart from current intellectual capacity, psychiatric formulation is concerned with alterations in cognition, which may be related to or constitute illness: deterioration after traumatic or other insults to the brain, degenerative disorders such as dementia, and schizophrenia, in which there is evidence of cognitive decline in many patients. The patient's most senior position in employment can be a good indicator of what the patient was capable of, alongside an indication of what the patient actually had to do. The ubiquitous nature of euphemisms in job description may disguise an appointment that in reality required little skill, responsibility or motivation. An informant history may be extremely useful in mapping out what has changed since then: people living with or closely acquainted with the patient are in a good position to notice this. They may be asked about, or spontaneously describe, absent-mindedness, forgetfulness, lack of concentration, vacant spells, lack of understanding of what is said, loss of interest in intellectual pursuits such as reading and keeping up with current affairs, and carelessness in completion of tasks. The natural history of the change should not be omitted: when did it start, were there any clear precipitants, and was it acute, insidious, stepwise or fluctuating?

Enquiries should always be made regarding drug and alcohol abuse and dependence. Although alcohol-induced cognitive impairment is recognized in several guises (delirium, persistent dementia, amnesic syndrome), the literature on the cognitive impairment produced by illicit drug use is in its infancy. Alternatively, the possibility of exposure to environmental or industrial toxins should not be disregarded.

COGNITIVE ASSESSMENT IN MENTAL STATE EXAMINATION: NOT JUST AT THE END

In the author's experience, junior doctors prefer to keep things simple. The problem with this approach is that it can obscure the integration of relevant observations into a coherent whole, which is actually the purpose of the psychiatric history, examination and formulation. Otherwise, a series of disconnected comments is produced, tick-box fashion, which fails to inform diagnosis or management.

The patient's appearance may be relevant to cognition. A classic example is dressing apraxia, in which the patient puts their clothes on the wrong part of their body. An unkempt, dirty and malodorous appearance is non-specific but may be associated with learning disability, dementia or schizophrenia, where self-care skills are poorly developed or lost. Similarly, in the assessment of behaviour, there are classic phenomena related to organic disorder, such as perseverative *utilization behaviour* seen in patients with frontal lobe disorders, in which objects are used repeatedly, for instance wearing three pairs of glasses on top of each other. A more common example is the *catastrophic reaction*, also known as *emotional incontinence*, in which the dementing patient suddenly becomes very angry or upset for relatively minor reasons.

Facial expression can give some clues as to intellectual status: again, patients with dementia may have a fixed or vacant expression, while patients with learning disabilities may appear placid, child-like or restricted in their facial and general emotional expressions. Emotionally inert patients may present so for a variety of reasons, including cognitive compromise. There is a recognized association, for instance, between the negative syndrome in schizophrenia, with its poverty of emotional responsiveness, and intellectual decline.

Much may be inferred from the patient's manner of speaking. Indeed, formal cognitive assessment may be entirely unnecessary in the patient who gives an articulate, fluent, detailed and accurate account, who clearly understands all that is asked of them, and who demonstrates normal memory and concentration in their conversation. It makes little sense to ask such a patient if they know where they are, who they are and what time it is at the end of the interview. By contrast, patients with any kind of general intellectual impairment may speak little, and in simple concrete terms, betraying poor general knowledge and awareness of themselves and their environment. Even if the patient says a lot, they may impart very little information. They may struggle to understand and to answer more complex enquiries, look puzzled or perplexed, fail to come up with expected information, and demonstrate a surprisingly poor memory for fundamental facts about themselves and their families during their account. This may extend to the names and ages of their partner and children, for instance. A classic example is age disorientation in schizophrenia,

when the patient says they are many years younger than their chronological age. This is considered cognitive rather than delusional.

Patients' language structure itself may be abnormal, as in expressive and receptive dysphasias, leaving aside formal thought disorder. Alternatively, poor educational standards are evinced by an overt ignorance of grammar (particularly the proper use of pronouns and verb conjugation in the author's experience) and correct pronunciation, although the vagaries of marked regional accents need to be taken into account to a certain degree.

Patients with delirium produce speech that is rambling, incoherent, disjointed and irrelevant. Acute intoxication with alcohol or other substances may present similarly. Such speech may be reminiscent of that of schizophrenic patients with marked thought disorder, but the latter should be fully conscious, alert and reasonably oriented, without the characteristic disturbance of sleep–wake cycle, fluctuation and worse presentation at night of the confused patient. Dysphasic patients produce characteristic speech anomalies: while the patient with receptive dysphasia may sound thought-disordered, with fluent but less than intelligible speech, expressive dysphasia manifests as very restricted speech and word-finding difficulties despite adequate comprehension.

Psychotic phenomena, by contrast, are linked less obviously to cognition. It is generally considered that the ability to experience and to express abnormal ideas and perceptions is related inversely to cognitive function: hence, the diagnosis of schizophrenia is difficult if not impossible to confirm in patients with moderate or severe learning disability. There are a handful of specific psychotic phenomena that may support diagnoses with cognitive impairment, such as visual hallucinations in Charles Bonnet syndrome and Lilliputian hallucinations in delirium tremens. Despite a great deal of research into the neuropsychological underpinnings of psychotic phenomena in schizophrenia throughout the 1990s, positive symptoms do not map easily on to the well recognized cognitive deficits manifest in more obviously organic brain disease, particularly deficits associated with circumscribed lesions.

Similarly, the relationship of insight to cognitive function is far from simple. Loss of insight is, like emotional inertia, a mental state anomaly that is possibly but not inevitably associated with cognitive impairment. In many patients, partial or absent insight may be understandably attributable to psychological defence mechanisms and personality factors. Do not forget to ask the patient whether they have noticed any problems with their memory or their ability to concentrate and pay attention. Can they take in a favourite TV programme, or read and remember a magazine or newspaper article? Are they absent-mindedly losing things, forgetting appointments, forgetting to switch off kitchen equipment, or becoming unaware of what people have said to them?

SPECIFIC COGNITIVE ASSESSMENT IN THE MENTAL STATE EXAMINATION

By the end of the mental state examination, it should be reasonably clear whether and what simple cognitive tests are appropriate. Essentially, these tests screen for clinically significant cognitive impairment in a variety of important intellectual functions. Language barriers should be taken into account in patients whose mother tongue is not English. The skill in appropriate cognitive assessment consists of deciding how much or how little to carry out, administering very simple bedside screening procedures for a reasonable spectrum of cognitive function where indicated, and adding further tasks if impairment is suggested by screening. Beyond this, the services of a clinical psychologist may be required.

As mentioned above, examination of orientation to place, person and time may be superfluous. Disorientation to time of day has the reputation of particular sensitivity, however. Ideally the patient should not have sight of a clock or a watch.

Use common sense in the interpretation of mistakes: patients living in a deprived environment may have little reason to know the exact date or even the day of the week – this is not the same as being presented with the date or day and then forgetting it. The patient should, however, know the month (except at the end or beginning of the month) and the year. If in doubt, ask additional questions such as what season it is, whether the patient has had breakfast today, and what sort of a job the patient does, in order to evaluate the patient's grasp and general awareness of their surroundings. Asking the patient to tell you what has been going on in the news can also be informative, particularly if momentous or highly publicized events have just taken place.

Memory should be divided into short-term or working memory (see below) and longer-term memory. Working memory can be tested by digit span – that is, repeating a series of digits after the doctor says them. The normal range is 5–9. Reversed digit span may also be tested and should normally extend to four or more digits. These and similar tests, such as saying the days of the week or the months of the year backwards, test attention and concentration as well as pure working memory. There is a good deal of overlap between the two. Mental arithmetic is another way of testing attention and concentration, plus the specific ability to calculate. Most patients find 'serial sevens' – subtracting seven repeatedly from 100 – rather challenging. Subtracting three from 20, or counting backwards from 20, is easier: again, take into account educational achievement.

Longer-term memory can be assessed by asking the patient to remember a name and an address comprising five or six elements. Use common sense: do not present something that is highly irregular or that you might not remember either. Make sure the patient has encoded the name and

address by asking them to repeat it after you, correctly, at least once. Then ask the patient to perform other brief tasks for 2–5 minutes before asking them to repeat the name and address. The patient should be able to remember all of it, especially if it was only five items long.

Semantic memory may be assessed by asking the patient some general-knowledge questions, such as the name of the current monarch and prime minister, and the years during which the Second World War took place. Some questions may be so easy as to exclude the effects of incomplete education and disinterest in current affairs.

Frontal lobe function was classically assessed in terms of abstract conceptual ability – that is, asking patients to think beyond the concrete, by explaining the meaning of proverbs, or explaining the differences or similarities between objects. Proverb interpretation is in the author's experience performed very poorly, particularly by medical students asked the meaning of, 'It's an ill wind that blows nobody any good'. Again, take into account educational and cultural factors and keep it simple: 'Too many cooks spoil the broth' is probably OK to start with. Differences such as that between a fence and a wall, a child and a dwarf, and a table and a chair may be followed by an enquiry as to how the items are similar. More difficult abstract pairs include a poem and a statue, and praise and punishment. Reasoning and judgement can be examined by asking the patient to 'guesstimate' how fast racehorses gallop, the height of the average woman in the UK, the age of the oldest person in the UK, and the best paid occupation in the UK.

Frontal function is a somewhat overinclusive entity, unfortunately. What this means is that abstraction is not sufficient on its own. The best recognized frontal functions that lend themselves to bedside testing, apart from abstraction, are initiation (i.e. mental response to an external stimulus), maintenance of set (i.e. inhibition of inappropriate responses) and set-shifting (i.e. changing response as appropriate). To test initiation, ask the patient to generate as many animals as they can think of in one minute: ten or fewer indicates impairment, as does perseveration (repeating an answer). Similarly, ask the patient to say words beginning with the letter 's' (or 'f' or 'a') in one minute; again, fewer than ten per minute and perseveration indicate definite impairment.

Frontal lobe tasks tend to be performed poorly by people with learning disability, dementia or schizophrenia.

In the patient in whom dementia or other organic psychosyndrome is considered, it may be well worth carrying out tests of more localized dysfunction (see below). The tests from the MMSE comprise a useful screen: naming a watch and a pen (to which may be added naming the hands, winder, nib, clip, etc.), writing a sentence, repeating a phrase, following the written instruction 'Close your eyes', following a three-stage spoken instruction, and copying intersecting pentagrams. To this may be added basic praxis, for example, asking the patient to demonstrate how they would comb their hair, strike a match and wave goodbye.

WHY NOT JUST DO A MINI-MENTAL STATE EXAMINATION?

There is no doubt that the MMSE is a very well established screen for clinically significant cognitive impairment. However, it is based almost entirely on memory, with virtually no frontal testing. If the patient scores less than 24 on the MMSE, then this merely raises the question of definite deficit, which will not have been addressed in more specific cognitive or diagnostic terms. The MMSE does not constitute an end in itself.

More importantly for junior doctors, in the Royal College of Psychiatrists' clinical examinations that evaluate the candidate's ability to assess cognition, the MMSE test sheet is not available. It is of course possible to administer the MMSE from memory, and lots of candidates do, but it is highly likely that a more specific evaluation will be required. Candidates facing standardized clinical stations with role-players may be given instructions to examine any area of neuropsychological function and not just the presence of clinically significant cognitive impairment as per MMSE score. Outside formal examinations, however, the MMSE is clinically useful when there is a question-mark over whether the patient is cognitively impaired. Nonetheless, this circumstance should not be the rule, given the many other avenues of exploration possible. The other use of the MMSE is as repeated measure, to track emerging cognitive decline or response to treatment, particularly in dementing illness where memory deficit is to the fore.

CLINICAL PSYCHOLOGISTS AND PSYCHOMETRIC TESTING

Historically, clinical psychologists 'followed diagnosis with numbers' – that is, carried out IQ tests, memory tests, language tests, etc. at the behest of the psychiatrist, whose provisional diagnosis implied an impairment. Clinical psychology has now become a diffuse discipline in which psychometric testing is very much a minority interest. A vague non-specific referral is likely to lead to a 'full assessment' and 'formulation', which may prove completely free of any neurocognitive content whatsoever. It therefore behoves the referrer to know what they are talking about, to have carried out a range of bedside tests, to know the basics of more formal evaluation, and to demonstrate justification for what is asked for. The request should state what neuropsychological deficit is suspected and why, and should indicate what specific psychometric tests may be useful.

Psychologists sometimes return reports devoid of numbers, for instance describing current IQ as 'normal' or 'average'. This tendency should be discouraged and precision

aimed for. After all, the investigation may need to be repeated in the future. The author would, furthermore, suggest that all psychiatrists should be capable of carrying out a premorbid IQ estimate, the National Adult Reading Test (NART)[5] and a current IQ estimate, the Quick Test.[6] The standard psychometric IQ evaluation, the Wechsler Adult Intelligence Scale (WAIS),[7] is a very lengthy test, onerous for both the patient and the clinical psychologist. Indeed, it is often 'pro-rated' – that is, a selection of sub-tests, rather than all of them, is administered and an IQ score derived from partial completion.

Finally, it is always well worth ordering all the patient's previous case notes. A summary of notes, although time-consuming, may generate much useful information about family history, development and previous personal function, which may not be available at initial interview. Occasionally there may have been a psychological assessment, even including an IQ score or other psychometric tests.

FROM THE CLINICAL TO THE THEORETICAL: STRUCTURAL–FUNCTIONAL RELATIONSHIPS

As a whole, cognitive function or the lack of it is as heterogeneous and variable as the spectrum of psychiatric disorder. Taking a hierarchical approach, the most gross cognitive impairments and anomalies are likely to be encountered in the dementias, learning disability and other organic psychosyndromes, produced by the usual range of neurodevelopmental, traumatic, neoplastic, metabolic, endocrine, vascular, infective, autoimmune, toxic and degenerative insults. An isolated area of cognitive dysfunction that can be attributed to a localized lesion, in the setting of preserved performance otherwise, is particularly likely to afford specific diagnostic implications. Moreover, 'functional' psychotic disorder, particularly schizophrenia, is increasingly recognized as including cognitive impairment in frontal function and long-term memory. These are disproportionate deficits considering overall intellectual ability. Even so, a fall in IQ from premorbid levels is extremely common in patients with schizophrenia. Similar chronic impairments are beginning to be described in bipolar affective disorder.

On the other hand, 'neurotic' disorder, personality disorder and substance (apart from alcohol) abuse were never of great interest in cognitive terms, although this is an emerging experimental field. Even so, it has long been recognized that depressed patients tend to complaints of absent-mindedness or inability to concentrate. On testing there can be a patchy impairment of memory and concentration owing to anxiety and difficulty in paying attention.

Occasionally, patients present with cognitive impairment that is more apparent than real, such as in dissociative amnesia and depressive pseudo-dementia. It pays to be aware of discrepancies between the pattern of impairment expected and that produced by the patient. Furthermore, there are particular issues with learning disability in the author's experience. Much subnormal IQ, whether original or recently acquired, goes undetected in psychiatric practice. Conversely, it is not rare to encounter patients thought to be learning disabled by other professions and who return a Quick Test estimate within one standard deviation of the mean.

Some important cognitive functions, principally attention, concentration, memory and executive function, are not localized to specific hemispheres or areas of the brain. These have a distributed neural basis and tend to be compromised by generalized cerebral dysfunction, by definition more than minor although not always permanent. Other functions such as verbal skills and visuospatial function are lateralized to the left and right hemispheres, respectively, depending on cerebral dominance. Discrete unilateral lesions, therefore, may have very different consequences compared with distributed compromise. Furthermore, the causes tend to differ: metabolic, electrolytic, infective, pharmaceutical and endocrine pathology is unlikely to target some parts of the brain more than others, as opposed to cerebrovascular, neoplastic and in some instances degenerative pathology.

The frontal lobes are something of an exception to the basic diffuse versus localized distinction, since they have a crucial integrative role across a number of important functions. Therefore, limited lesions of the frontal lobes can impact upon distributed functions. This is probably why 'frontal lobe function' is such an inherently overinclusive and unsatisfactory term.

DISTRIBUTED FUNCTIONS

Attention and concentration

The reticular activating system connects the brainstem to the thalamus and neocortex via ascending cholinergic, monoaminergic and serotonergic pathways. There is neocortical feedback, particularly from the prefrontal cortex and the posterior parietal and inferior temporal cortex, all of which receive afferent perceptual input. The result is conscious attention. Disruption of attention leads to confusion, termed delirium at the moment, at least when it comprises an acute presentation attributable to a new insult to the brain rather than a chronic degenerative picture.

Like any other disorder, there is a spectrum of severity applicable to attentional impairment. Clinically, it presents as poor orientation, an inability to sustain attention, distractibility by irrelevant stimuli, and inability to inhibit inappropriate responses. Such patients may perseverate, fail to repeat and reverse strings of digits, fail to reverse the days of the week, fail simple mental arithmetic, and fail to learn a name and address. There is obviously an overlap

with compromise of working memory. Although usually the result of a diffuse insult, right-sided prefrontal lobe lesions are supposedly more likely to produce attentional impairment than the left-sided variety. Many delirious patients are quiet and lethargic, but some are actively agitated, deluded and hallucinated: this phenomenon is recognized in severely physically ill patients as 'ICU psychosis'. Some forms of delirium, the classic example being delirium tremens, are usually accompanied by hallucinosis.

Memory

The overall concept of memory is neuropsychologically quite complex and occasionally rather bewildering. The major subclassification into implicit and explicit memory has much more theoretical than clinical utility. Essentially, explicit (also known as declarative) memory subsumes what can be recalled by conscious (i.e. voluntary) recollection. Implicit memory is something of a 'rag-bag' and includes everything else. First, implicit memory includes what the patient has learned (i.e. not forgotten) but cannot voluntarily remember on demand. Such implicit memory material can be brought to consciousness by priming or cueing: if the patient is given a list of ten words to learn but can recollect only five of them, they may recall some of the others if given the initial letter or other clue. Second, implicit memory includes the learning and retention of motor skills, such as driving a car, typing, or playing a musical instrument. This is known as procedural memory. Third, implicit memory subsumes other non-verbal but entirely unconscious forms of learning such as conditioned reflexes.

Implicit memory tends to be relatively preserved, even in serious psychotic or degenerative disorder, and is not assessed routinely in psychiatry. However, loss of motor skills, if the patient had any, can be asked about in the history, particularly if an informant is available. It should not be forgotten that motor skills may deteriorate without regular practice, leaving aside the effects of mental disorder.

Explicit memory is divided into short-term memory (also known as working memory) and long-term memory. The working memory system serves to keep small amounts (five to nine items) of verbal, auditory or visual material present in conscious awareness for recall or manipulation, or both, for a brief period of up to 30 s. There is an obvious overlap with conscious attention. Working memory functions alongside long-term memory, but surprisingly the acquisition and retrieval of long-term memories can occur despite defective working memory.

Long-term memory is divided into episodic and semantic memory. Episodic memory is what most people would call memory – the recollection of personal experience – for instance, what one did yesterday. Semantic memory is what most people would call knowledge: the capital of France is Paris, and a dog is a four-legged domesticated animal that barks and wags its tail.

An added layer of complication in the understanding of memory is the processes through which the systems work. In other words, information of whatever kind must be encountered (consciously or otherwise), encoded, reinforced or consolidated, and then accessed for recall; otherwise, the information cannot become a memory. This may have practical clinical implications: for instance, the author was asked to complete a second opinion on a patient whom several 'assessments' had concluded was suffering from post-traumatic stress disorder. The flaw in the argument was that the traumatic incident comprised a blow to the head that produced unconsciousness accompanied by several minutes' amnesia both before and after the event. Logically, the patient could not suffer re-experiencing and avoidance symptoms related directly to an event of which he had no memory at all. This elementary error had understandably compromised the patient's management for the 3 years since presentation.

Another relevant process is forgetting: there are further theoretical distinctions between the functions of searching or accessing memories, and the existence of a store where representations are held. Recall is thought to be harder than recognition for these reasons: recall of material requires a search for and access to what is in the store, while recognition involves only a check for familiarity between the material presented and the content of the store.

Although memory is overall a distributed function, its subdivisions do have localized affiliations. Thus, the basal ganglia and cerebellum are thought to be of particular importance in conditioning and procedural memory. Regarding working memory, the substrates for verbal content (phonological loop) and non-verbal content (visuospatial sketch pad) devolve to the dominant and non-dominant hemisphere, respectively, while the frontal cortex includes a central executive function to operate them effectively. The circuit of Papez, a component of the limbic system comprising the medial temporal lobe, diencephalon, basal forebrain nuclei, fornix and cingulate gyrus, is essential for episodic memory. Semantic memory is lost with lesions of the dominant temporal neocortex.

Classic neuropsychological distinctions between these types of memory are of limited utility in psychiatry. Clinically, the ability to acquire, retain and recall new episodic material (i.e. anterograde memory) as opposed to the ability to recall old episodic material (retrograde amnesia) is much more relevant. Working memory and attention are usually compromised in anterograde amnesia, alongside this long-term episodic memory failure. The Rivermead Behavioural Memory Test (RBMT)[8] is the standard measure of episodic memory function in anterograde terms. The RBMT includes both non-verbal and verbal elements (words, drawings, faces) and tests of both recall and recognition. Interestingly, patients with schizophrenia perform very poorly on this test and on tests of semantic memory. The bedside test of verbal episodic memory recall, asking a name and address learned a few minutes previously, is a useful screen. Retrograde memory evaluation turns upon

whether the autobiographical history given by the patient can be validated. In some patients it may be obviously confabulatory rather than missing. Formal tests exist, such as the Famous Faces Test, in which the patient is presented with photographs of famous people or events from different periods in the past.[9,10]

Amnesic syndrome is an important entity in psychiatry: anterograde or retrograde amnesia, or both, occur in the absence of other cognitive dysfunctions, including, surprisingly, implicit memory. Amnesic syndrome can be acute or chronic.

Acute amnesia occurs in delirium and in transient global amnesia: it may present with alcohol and drug intoxication, closed head injury and focal epilepsy. Anterograde amnesia is the rule, alongside some working memory/attentional impairment. Patients cannot remember anything for more than a few minutes: attentional impairment inevitably results in a degree of disorientation in place and time. Nonetheless, in transient global amnesia, the patient may have normal working memory and attention (i.e. normal digit span and reversal tests) and be acutely aware of their difficulty, repeatedly asking where they are, how they arrived at that location and what has happened. Furthermore, in transient global amnesia, there is a dense retrograde amnesia that may go back for decades. The patient may be unaware of the death of a relative or the achievements of their children, for instance.

Chronic amnesic syndrome indicates long-term bilateral damage to the medial temporal lobe, particularly the hippocampus, as in, for example, early Alzheimer's disease, herpes encephalitis and posterior cerebral artery stroke. It is also associated with closed head injury. Diencephalic damage is a classic complication of alcoholism in the form of Korsakoff psychosis. In chronic amnesic syndrome, there is a preserved intelligence quotient but severe anterograde amnesia accompanied by retrograde amnesia. This is much worse with diencephalic than hippocampal compromise. There is a temporal gradient that is counterintuitive, in that more distant memories are clearer than more recent memories. Working memory, attention and aspects of implicit memory are unaffected.

Semantic memory disorder occurs in Pick's disease, herpes encephalitis and severe schizophrenia. It also surfaces in progressive dementing disorder, particularly Alzheimer's disease. Bedside tests of semantic memory are not usually indicated, but there is some overlap with the 'frontal' category fluency test (i.e. naming as many items as possible, such as animals, or words beginning with a specific initial, in a brief period such as 1 minute). Asking about general knowledge (e.g. name of prime minister) is also useful. The patient may be asked to verify (or otherwise) statements such as 'Admirals are in charge of ships' and 'Beer is sold at a butcher's shop'. By far the best formal test of semantic memory is the Hodges semantic memory battery,[11] which has been used widely in formal neuropsychological assessment and research. The Graded Naming Test is much shorter: the patient is asked to look at a number of pictures of increasingly obscure items and say what they are.[12] The Silly Sentences Test consists of correct and nonsensical statements as above, the patient being asked to determined which is which.[13]

Frontal lobe function

Frontal lobe function, also known as executive function, is problematic in terms of its heterogeneity and ubiquity. In summary, frontal lobe function involves generating responses to stimuli, choosing the most appropriate, inhibiting inappropriate responses, planning and problem-solving, keeping a sequence of behaviours on course to achieve an overall goal, and yet changing response parameters when required by environmental variation. Bedside tests for frontal function have arisen from the study of patients with circumscribed frontal lobe lesions.

Frontal lobe compromise can be suggested by the history. Disinhibition in mood and behaviour and a lack of empathy for others are of course not specific to patients with frontal lobe disorders, but they can be of sufficient magnitude and difference compared with the normal personality and behaviour to be significant. Lack of suppression of inappropriate internal stimuli is thought to be responsible. On the other hand, the apathy and marked slowing also observed in frontal patients reflects that they may respond very poorly to internal and external stimuli. Adequate response to external stimuli enables patients to attempt testing, although this must be done carefully if cooperation is not to be lost. Category fluency for animals (or vehicles or household items, for instance) and words beginning with specific letters can be considered a frontal test. Perseveration occurs when items are repeated (the patient gets stuck rather than forgets they have said the item or word already), and irrelevant items or words may be offered owing to distractibility; one schizophrenic patient of the author's responded to a request to name sea creatures with the phrase 'men in aqualungs', for instance. Planning and strategy is compromised: most people when asked to name animals would use superordinate categorization to identify groups of animals, such as thinking of farm animals, domestic animals or jungle animals. Patients with frontal lobe disorders do not do this as effectively.

If frontal impairment is suspected, it may be useful to ask the patient to carry out non-verbal tasks as well as proverb, difference and similarity explanation, cognitive estimates and category fluency (see above). The patient may be asked to copy an alternating sequence of triangles and squares (Figure 7.1) or to tap on the desk twice when the examiner taps once (under the table), and vice versa (i.e. the patient should tap once when the examiner taps twice). This is repeated until the examiner is satisfied that the patient can or cannot do it. Further formal testing is within the remit of clinical psychologists. They may use a structured initiation task such as the formal FAS fluency test[14] (the patient is

asked to say as many words as they can think of beginning with the letter F in 1 minute, then with the letter A, and then with the letter S) and 'dual tasks' of divided attention (since the working memory central executive is compromised), such as the Stroop Test[15] and the Trail Making Test.[16] Divided attention clearly overlaps with the ability to maintain and change set. The standard set maintenance and shifting task is, however, the Wisconsin Card Sorting Test, in which cards are sorted according to the different numbers of various shapes of separate colours that appear on each one.[17] The Tower of London (or Hanoi) test is used to assess planning and sequencing.[18] The Cognitive Estimates Test comprises ten 'guesstimate' questions (see above) and assesses reasoning and judgement.[19,20] The Behavioural Assessment of the Dysexecutive Syndrome (BADS)[21,22] is an ecologically valid test that covers functional aspects of frontal impairment.

Causes of frontal lobe syndrome include focal dementia such as Pick's disease, anterior cerebral artery stroke, head injury (closed or local trauma) and tumours. Frontal signs are eventually seen in degenerative disorders that have the effect of isolating the frontal lobes, for example subcortical dementias such as Huntingdon's and Parkinson's diseases, progressive supranuclear palsy, leucodystrophies and acquired immunodeficiency syndrome (AIDS) encephalopathy.

LOCALIZED COGNITIVE FUNCTIONS

The dominant hemisphere

The left hemisphere is dominant in right-handed people and about half of left-handed people: its functions include language, calculation and higher motor control (praxis).

Dysphasia (impairment of language function) should be distinguished from dysarthria (failure of articulation). Dysarthria has a variety of causes, the most common in psychiatric practice probably being antipsychotic drug treatment with extrapyramidal side effects. Cranial nerve palsies, cerebellar and basal ganglia pathology, and anterior left hemisphere lesions can also contribute to dysarthria.

Dysphasia presents in two forms, receptive and expressive. In receptive dysphasia there is a lesion affecting Wernicke's area in the dominant posterior superior temporal lobe. Language is not understood and speech output is fluent but unintelligible to varying degrees, including neologisms, grammatical and syntactical errors, and overinclusiveness. In expressive dysphasia there is a lesion affecting Broca's area in the dominant inferior frontal lobe, usually extending into the frontoparietal region served by the superior middle cerebral artery. Language is understood but the patient cannot find words and assemble phrases and sentences: speech output is simple, minimal and not entirely correct in structure, but it betrays better comprehension than the irrelevant speech output of patients with expressive dysphasia. Both types of dysphasia affect the ability to repeat and to name objects. The Graded Naming Test can be used to evaluate naming impairment and semantic memory. Repetition can be examined simply using the MMSE phrase 'No ifs, ands or buts'. Language comprehension can be evaluated with the Token Test: a number of tokens of different shapes and colours are assembled before the patient, who is then given standard verbal instructions of variable complexity, such as 'Put the small red circle on the big green square'.[23] Furthermore, most dysphasic patients are dyslexic, having difficulty in reading aloud or comprehension of text: they can be asked to read a short passage from a book, newspaper or magazine, and then describe its content. The distinction should be drawn between patients with learning disability and illiterate patients, however, and those who have lost a pre-existing ability. Similarly, there may be dysgraphia (difficulty in writing): the patient may fail to write a sentence or copy text. Dominant parietal lobe damage may be associated with dyspraxic dysgraphia, in which the motor control of writing is impaired. Both dysgraphia and dyslexia may present as neglect phenomena owing to non-dominant lesions (see below): the left (first) parts or words are not written or not read at all, or they are read or written incorrectly.

To the rear of Wernicke's area is the angular gyrus, which connects the temporal, parietal and occipital lobes. Lesions here produce Gerstmann syndrome: dysgraphia, dyscalculia, right–left disorientation and inability to name individual fingers according to convention (finger agnosia). The angular gyrus appears very important for numeracy in general. If a patient cannot carry out serial subtractions of three from 20, then simpler arithmetic function can be checked, for instance single-figure addition. If further examination is required, the patient may be asked to point to numbers on command, copy numbers, write numbers to dictation, and read aloud written numbers. Again, the context of overall intelligence quotient and educational and motivational factors should be considered.

Finally, apraxia, the loss of simple motor skills, can be a consequence of dominant hemisphere compromise. The patient may be asked to demonstrate buccofacial actions (how they would lick lips, cough, sip through a straw, blow out a match), limb actions (how they would wave goodbye, beckon 'come here', salute, hitch a lift) and object usage

Figure 7.1 Ask the patient to copy the connected alternating shapes above

(how they would brush teeth, comb hair, strike a match, use scissors, hammer a nail). If there is difficulty, the action can be demonstrated by the examiner to see whether the patient can copy it.

The non-dominant hemisphere

Localized non-dominant hemisphere lesions result in neglect (lack of awareness) phenomena and visuospatial defect, particularly lack of visual recognition. Stroke-induced lesions can result in denial that the affected side exists, denial of hemiplegia (anosagnosia) or minimization (anosodiaphoria). Patients may ignore visual, auditory and tactile sensory input, usually from the left side of the body, since the right hemisphere is generally non-dominant. Extinction occurs when the patient is aware of such stimulation but, if the stimulus is presented bilaterally, ignores the stimulus on the neglected side. Neglect of space outside the body can occur after damage to either hemisphere but is persistent and severe following non-dominant insults. This can be determined by asking the patient to copy a clock face: half the numbers are left out or they are all squashed into one side. Other drawings are copied with the left half missing. Another aspect of visuospatial difficulty is manifest as dressing apraxia.

Visuospatial function can be tested at the bedside by asking the patient to copy simple line drawings such as the intersecting pentagons in the MMSE or a three-dimensional drawing of a cube. The Visual Object Space Perception Battery[24] is a very useful test for further evaluation of visuospatial function, although some of the more difficult tasks have a distinct executive element.

Visual agnosia (inability to name objects presented in pictorial form) may result from semantic memory impairment or deficits in visual processing (apperceptive agnosia). Patients with apperceptive agnosia cannot copy line drawings or match identical depictions. However, they may be able to identify an object by touch. The patient should be able to describe the attributes of an object named by the examiner, unlike patients with semantic memory disorder. Apperceptive agnosia may extend to an inability to recognize faces (prosopagnosia). Even so, patients with prosopagnosia may recognize people from their voice, gait or dress and can match the same face from portraits or photographs. They can describe facial features and expressions. The Recognition Memory Test[25] is a formal evaluation including faces presented singly and then in pairs, the patient needing to identify the face seen previously.

ESTIMATING PREMORBID AND CURRENT INTELLIGENCE QUOTIENT

Given the impact of current IQ upon personal and social function, it is important to judge the patient's general intellectual capacity to inform formulation and management. The population mean IQ score is 100: the range is between zero and 200. The standard deviation is 15. People with an IQ outwith two standard deviations of the mean are statistically abnormal: hence, learning disability is defined by an IQ below 70, while an IQ of over 130 denotes superiority. The measurement error is five points. For this reason, a person may be classed as having a learning disability with an IQ of 75 if there are sufficient impairments in (separately assessed) adaptive functioning. Conversely, if there is adequate adaptive functioning, then learning disability would not be diagnosed in the context of an IQ of 65. The low score would be attributed instead to measurement error. Because practice effects are observed in neuropsychological evaluation, and the WAIS does not include multiple versions, it can be misleading to repeat the evaluation particularly over a brief period of time.

WAIS assessment is, as mentioned above, rather lengthy and cumbersome, potentially presenting the individual with multiple experiences of failure, which can be demoralizing. Furthermore, IQ evaluation may be seen as undesirable for social reasons, undermining the value of people and constituting grounds for discrimination. IQ evaluation measures, it has been argued, no more than the ability to complete IQ tests successfully. Nonetheless, in cognitive terms, IQ tests load quite heavily on aspects of frontal and memory function, which are know to be crucial for independent living skills.

Availability of the WAIS equipment is limited by its publisher to properly qualified clinical psychologists. A good compromise is the Quick Test (Figure 7.2; Table 7.1).

Premorbid intelligence is of great interest where IQ decline is suspected, as in degenerative or other instances of permanent brain damage however caused. Estimation requires a test that accesses aspects of cognition at least relatively preserved subsequent to psychiatric or neurological disorder: scores on the test should correlate highly with current IQ in the normal population. The best established premorbid IQ evaluation is the NART (Table 7.2). This test asks the patient (whose first language should be English) to read 50 words whose spelling does not reflect their pronunciation. The implication is that because reading one's mother tongue is such a well established skill, acquired in childhood, it is resistant to the effects of cerebral pathology.

IQ decline can be assessed by comparing the NART score with the Quick Test score. Substantial discrepancies are manifest in dementia, after serious insults with permanent effects, and in some patients with chronic schizophrenia. Both the presence of decline and the current IQ estimate can be useful information in terms of diagnosis and prognosis.

CONCLUSION

Every competent psychiatrist must have an adequate working knowledge of cognition: the generality and specifics of

Table 7.1 Quick Test. Reproduced with permission from Ref. 6

Word	Difficulty/mental age (years)	Answer	Score out of 49	IQ estimate	Word	Difficulty/mental age (years)	Answer
Belt	Easy	4	49	135	Graceful	10	1
Dancing	Easy	1	48	130	Fluid	11	2
Traffic	Easy	4	47	120	Solution	11	2
Whistle	Easy	4	46	116	Discipline	12	4
Fence	Easy	3	45	110	Crystallised	13	2
Drink	Easy	2	44	108	Turntable	13	1
Wreck	Easy	3	43	104	Saccharin	14	2
Music	Easy	1	42	102	Immature	14	4
Medicine	Easy	2	41	100	Cordiality	15	1
Gun	Easy	4	40	98	Velocity	15	3
Pepper	Easy	2	39	96	Decisive	16	4
Racing	Easy	3	38	92	Laceration	16	3
Salt	Easy	2	36	90	Foliage	17	3
Woman	Easy	1	35	87	Imperative	17	4
Sugar	Easy	2	32	84	Intimacy	18	1
Track	Easy	3	30	80	Concoction	18	2
School	6	4	27	75	Conviviality	18	1
Partner	7	1	25	70	Chevrons	18+	4
Couples	7	1	22	65	Condiment	Hard	2
Rail	7	3	21	60	Cacophony	Hard	3
Respected	8	4	19	55	Miscible	Hard	2
Betting	8	3	18	50	Imbibe	Hard	2
Daring	9	3	16	45	Amicable	Hard	1
Stadium	9	3	15	40	Pungent	Hard	2
Pedestrian	10	4					

Table 7.2 The National Adult Reading Test (NART). Reproduced with permission from Ref. 5

Words			Number of errors	Premorbid IQ estimate	Number of errors	Premorbid IQ estimate	Number of errors	Premorbid IQ estimate
Chord	Placebo	Assignate	0	131	17	110	34	89
Depot	Detente	Subtle	1	129	18	108	35	87
Bouquet	Puerperal	Gist	2	128	19	107	36	86
Capon	Gauche	Superfluous	3	127	20	106	37	85
Nausea	Leviathan	Banal	4	126	21	105	38	84
Courteous	Prelate	Cellist	5	124	22	103	39	82
Equivocal	Demesne	Zealot	6	123	23	102	40	81
Catacomb	Labile	Aeon	7	122	24	101	41	80
Thyme	Ache	Abstemious	8	121	25	100	42	79
Radix	Aisle	Idyll	9	120	26	98	43	77
Hiatus	Psalm	Aver	10	118	27	97	44	76
Procreate	Deny	Topiary	11	117	28	96	45	75
Gouge	Debt	Beatify	12	116	29	95	46	74
Simile	Rarefy	Sidereal	13	115	30	94	47	73
Quadruped	Naive	Syncope	14	113	31	92	48	71
Façade	Gaoled	Campanile	15	112	32	91	49	70
Drachm	Heir		16	111	33	90		

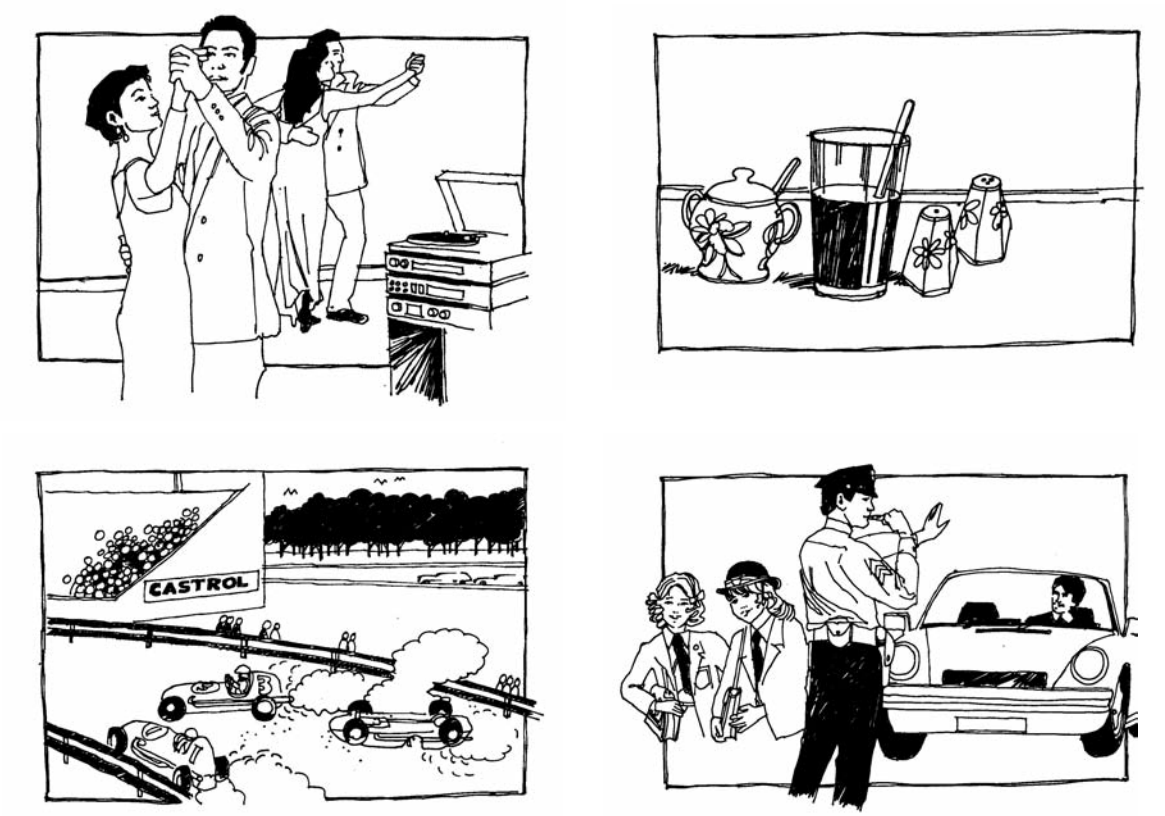

Figure 7.2 Quick Test. Place the cards in front of the patient on a table in a block of four. Leave a gap of about 15 cm between each card to make it easier for the patient to point at the picture they select in response to each test word. Reproduced with permission from Ref. 6

normal function, how it can go wrong, the clinical discovery of deficits, and their diagnostic and prognostic implications. Cognitive evaluation is, however, predicated not upon an afterthought of mental state examination but on foregoing referral information, and the patient's previous history, history at interview and mental state assessment. Only then can appropriate screening take place, followed up with further enquiry and referral to colleagues in clinical psychology when necessary.

Repeated and extensive practice with a broad spectrum of patients will illustrate how cognitive issues insinuate themselves into psychiatric formulation and so inform investigation, management and outcome. It is hoped that the material presented in this chapter constitutes a good starting point.

ACKNOWLEDGEMENT

The author is indebted to Professor John R Hodges, whose standard work, *Cognitive Assessment for Clinicians*,[26] very much informed a substantial minority of what is presented here. Professor Hodges' work is unreservedly recommended for further reading.

REFERENCES

1. Folstein M, Folstein S, McHugh P (1975) 'Mini-mental state': a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatry Research* **12**: 189–98.
2. Phillips C, Hon Y, Smith B, Sutton A (1986) Severe mental retardation in children from socially disadvantaged families. *Child: Care, Health and Development* **12**: 69–91.
3. Rantakallio P (1987) Social class differences in mental retardation and subnormality. *Scandinavian Journal of Social Medicine* **15**: 63–6.
4. Buchanan A, Oliver J (1977) Abuse and neglect as a cause of mental retardation: a study of 140 children admitted to subnormality hospitals in Wiltshire. *British Journal of Psychiatry* **131**: 458–67.
5. Nelson H (1982) *The National Adult Reading Test (NART): Test Manual*. Windsor: NFER-Nelson.
6. Bowen K, Mortimer A, Rooke-Ley S (1999). Measuring IQ in schizophrenia: the 'Quick Test'. *Cognitive Neuropsychiatry* **4**: 81–8.
7. Wechsler D. (1981) *Manual for the Wechsler Adult Intelligence Scale: Revised*. New York: Psychological Corporation.

8. Wilson B, Cockburn J, Baddeley A (1985) *The Rivermead Behavioural Memory Test*. Reading: Thames Valley Test Corporation.
9. Hodges J, Ward C (1989) Observations during transient global amnesia: a behavioural and neuropsychological study of five cases. *Brain* **112**: 595–620.
10. Estevez-Gonzalez A, Garcia-Sanchez C, Boltes A *et al.* (2004) Semantic knowledge of famous people in mild cognitive impairment and progression to Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders* **17**: 188–95.
11. Hodges J, Salmon D, Butters N (1992) Semantic memory impairment in Alzheimer's disease: failure of access or degraded knowledge? *Neuropsychologia* **30**: 301–14.
12. McKenna P, Warrington E (1983) *The Graded Naming Test*. Oxford: Cambridge Cognition.
13. Collins A, Quillian M (1969) Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behaviour* **8**: 240–47.
14. Benton A (1968) Differential behavioural effects in frontal lobe disease. *Neuropsychologia* **6**: 53–60.
15. Golden CJ (1978) *Stroop Color and Word Test: A Manual for Clinical and Experimental Uses*. Chicago, IL: Skoelting.
16. Reitan R (1958) Validity of the Trailmaking Test as an indication of organic brain damage. *Perceptual Motor Skills* **8**: 271–6.
17. Nelson H (1976) A modified card sorting test sensitive to frontal deficits. *Cortex* **12**: 313–24.
18. Shallice T (1982) Specific impairments of planning. In DE Broadbent, L Weiskrantz (eds) *The Neuropsychology of Cognitive Function*. London: The Royal Society.
19. Shallice T, Evans M (1978) The involvement of the frontal lobes in cognitive estimation. *Cortex* **14**: 294–303.
20. Nathan J, Wilkinson D, Stammers S, Low J (2001) The role of tests of frontal executive function in the detection of mild dementia. *International Journal of Geriatric Psychiatry* **16**: 18–26.
21. Wilson B, Evans J, Alderman N, Burgess P, Emslie H (1997) The Behavioural Assessment of the Dysexecutive Syndrome (BADS): behavioural assessment of the dysexecutive syndrome. In P Rabbit (ed.) *Theory and Methodology of Frontal and Executive Function*. Hove: Psychology Press.
22. Katz N, Tadmor I, Felzen B, Hartman-Maeir A (2007) The Behavioural Assessment of the Dysexecutive Syndrome (BADS) in schizophrenia and its relation to functional outcomes. *Neuropsychological Rehabilitation* **19**: 192–205.
23. McNeil A, Prescot T (2008) *The Revised Token Test*. Baltimore, MD: University Park Press.
24. Warrington E, James M (1991) *The Visual Object Space Perception Battery (VOSP)*. Bury St Edmunds: Thames Valley Test Company.
25. Warrington E (2008) *The Recognition Memory Test*. Windsor: NFER-Nelson.
26. Hodges J (2007) *Cognitive Assessment for Clinicians*, 2nd edn. Oxford: Oxford University Press.

PART 2

Developmental, behavioural and sociocultural psychiatry

This page intentionally left blank

Chapter 8

Human development

Marcus Munafò and Angela Attwood

CONCEPTUAL FRAMEWORKS

Nature and nurture

Are individual differences in behavioural and physical traits determined by genetic variation or by differences in environmental circumstances? Historically, theorists have tended to advocate extreme positions in this debate, but it is now widely accepted that genetic and environmental influences do not operate independently but are interactive. In most cases, genetic factors provide a set of parameters in which an infant can develop, with the exact nature of this development being determined by the environment. Thus, it seems that the important question is not *which* influence is most important, but rather *how* genetic and environmental factors interact to determine human development.

The heritability statistic used to express the relative contribution of genotype and environment to the observed behaviour or trait (i.e. phenotype) is widely misunderstood. This is a population statistic that provides information on the proportion of *variability* in the observed phenotype that may be accounted for by variability in the genotype (and, therefore, information on variability accounted for by the environment). The number of fingers on one hand, for example, although determined genetically, is not highly heritable, as any *variation* in this number is generally the result of environmental factors. Height, however, is highly heritable, since most diets – even relatively poor ones – are sufficient to promote growth, so that environmental variation plays a smaller role, with any residual variation being the result of genetic factors. The latter example also indicates that a heritability statistic for a given phenotype is specific to the time and population on which it was calculated: in populations where the diet is more variable, more of the variation in height in that population may be associated with this environmental factor, so that the heritability coefficient will be different even if the underlying distribution of genes is the same.

Stage theories

Many abilities and behaviours develop in a uniform way. The prerequisites for these behaviours, such as basic motor skills, are inherited, but through dynamic interaction with the environment children learn to develop these skills. Stage theories propose that development passes sequentially through qualitatively different phases. In most cases, accomplishment of one stage is necessary before development of the next can proceed. The rate of development is influenced by environmental factors, but some stage theories identify critical periods by which accomplishment of certain skills should be achieved.

Stage theories have been used to describe many aspects of development, including linguistic, moral and motor development. Piaget's theory of cognitive development is one of the best-known examples. The goal of stage theorists is to determine the distinct qualitative features of the developmental process and the important changes that characterize transition from one stage to the next.

Maturational tasks

Development of a skill or behaviour occurs when a child performs or practises tasks that strengthen, extend or consolidate current knowledge or abilities, so that higher levels of that skill or behaviour can be achieved. For example, a child will master the art of standing unaided before attempting his or her first steps, which will be unsteady at first, becoming more assured and steady over time. These tasks propel the child through the critical periods of development.

Maturation refers to a largely genetically determined 'blueprint' that sets out a child's critical periods of development (i.e. the sequential aspects of development that are under biological control). The concept of maturation is linked closely to the nature–nurture distinction, as it is the interaction between maturation and experience that shapes an individual's development.

Although it is relatively easy to understand maturation in the development of physical abilities, the applicability of this framework to psychological development is not accepted universally. It is also important to distinguish between simple maturation, which is the pure biological unfolding driven primarily by genetic factors, and development, which is the sequence of changes over the lifespan that results from the interaction of some behaviours and the environment.

Maturity

Maturity in a physical sense refers to the time in life when physical growth has ceased. From a psychological viewpoint, however, maturity is somewhat more difficult to define. Unlike growing in height, for example, there is no clear cessation in psychological growth. In fact, learning and development continue on an emotional, cognitive and intellectual level throughout an individual's lifetime. Therefore, it may be more relevant to think of maturity as the conclusion to a biologically predetermined pattern of development, or the culmination in the maturation process described above.

James Tanner devised a series of longitudinal charts to illustrate growth references on physical parameters, including height, weight, skin folds and puberty, ranging from birth to maturity. Their purpose was to provide a standardized tool for clinicians to assess whether a child's size and growth rate were within the normal limits for his or her age, sex, socioeconomic position and ethnic group, and to help diagnose growth disorders such as short stature and low growth velocity. Despite originally being designed for clinical use, the charts have also been used to predict adult physical characteristics such as height in adulthood. Questions have been raised regarding the validity of the charts, particularly due to trends of increasing size over time, which have been answered, to some extent, by revisions to the original work. There does, however, remain some confusion over the comparability of Tanner scales with other growth references. Tanner's charts are still in clinical use in the UK, often in combination with other growth references (Figure 8.1).

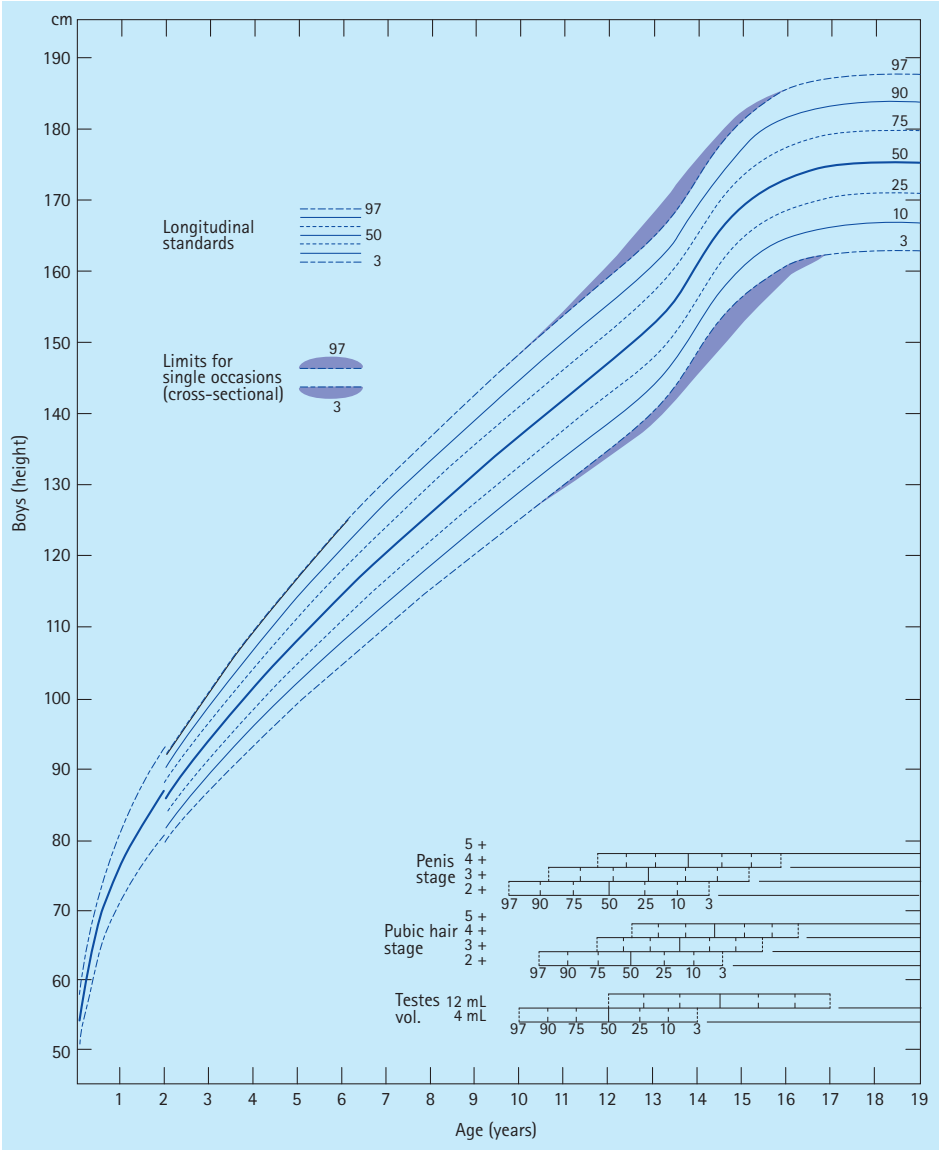

Figure 8.1 Example of Tanner growth chart for height. A series of charts of longitudinal growth standards for physical development (including parameters for height, weight, skin folds) were developed for children according to age, sex and sexual maturity. From Tanner JM, Whitehouse RH (1976) Clinical longitudinal standards for height, weight, height velocity, weight velocity, and stages of puberty. *Archives of Disease in Childhood* **51**: 170–79.

Gene–environment interactions

The broad implication of interactions between genetic and environmental influences has been described above. More recently, more specific findings have contributed to the debate concerning the nature and development of intelligence and personality, both of which have been claimed to be up to 80 per cent heritable, although most estimates range from 50 per cent to 60 per cent. Although twin studies and other methods have certainly suggested a genetic component to intellectual ability, the interpretation of these results is problematic, given the highly interactive nature of the factors in question and the complex nature of intelligence itself.

Early, simple additive genetic models of intelligence suggested that genotype and environment both contributed independently to the phenotypic expression of intelligence, so that the relative contribution could be determined by certain empirical means (e.g. the comparison of identical twins raised apart, or the comparison of monozygotic (MZ) versus dizygotic (DZ) twins). Notwithstanding the methodological limitations of many of these studies (and a few celebrated scientific frauds), the model of genetic influence guiding this research is potentially inappropriate. The model of genotype–phenotype interaction, proposed by Scarr and McCartney in 1983, highlights this (Figure 8.2).

Although genotype and environment will influence the phenotype of the individual, these relationships may not be unidirectional. Most importantly, the phenotype of the individual will influence the environment, so that physically short individuals will be less likely to find themselves playing basketball, for example. In this sense, it would be possible to find a genetic component for basketball-playing, but this would not be interpreted as a gene 'for' basketball (whereas comparable findings in intelligence are so interpreted). Rather, the combined influence of a number of genes related to physical development, when expressed in a given context (i.e. a population where basketball is played) would influence the likelihood of an individual playing basketball. The various influences in this interactive model include the child's genotype on the child's phenotype, the child's environment on the child's phenotype, the child's phenotype on the child's environment, the parents' genotype on the child's genotype, the parents' genotype on the child's environment, and various external influences on the child's environment.

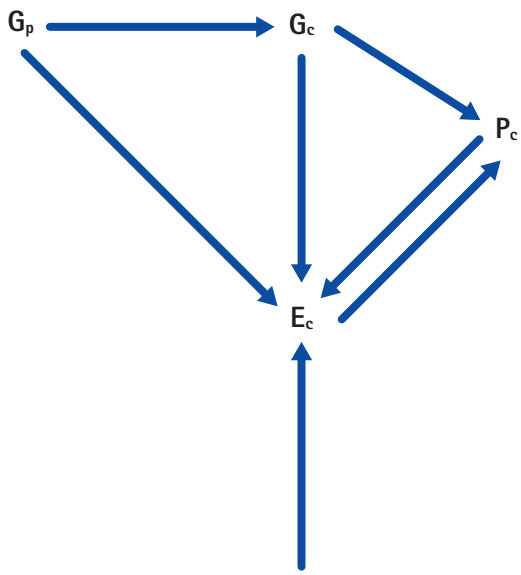

Figure 8.2 Scarr and McCartney's model of genotype–environment interaction in behavioural development. The child's observable behaviour (phenotype) will arise from an interaction between his or her genotype and rearing environment. The genotype of the parent will determine the genotype of the child and thus also indirectly influence the child's phenotype. Parental genotypes will have an additional influence over the child's rearing environment. The model also allows for a bidirectional influence between the child's behaviour and rearing environment, highlighting that the child actively contributes to their own behavioural development. E_c, child's rearing environment; G_c, child's genotype; G_p, parent's genotype; P_c, child's phenotype. From Scarr S, McCartney K (1983) How people make their own environments: a theory of genotype–environment effects. *Child Development* **54**: 424–35.

The influence of the parents' genotype on the child's environment may also, because of the obvious genetic relationship between parents and child, appear as a genetic influence rather than an environmental one. Finally, because of certain cultural norms, certain physical characteristics will result in the individual being treated in a certain way. Attractive children, for example, are likely to receive more stimulation in infancy, which may result in faster or improved cognitive development. This may appear as a genetic influence, even though the gene found is 'for' physical attractiveness, which, in a specific culture, results in certain treatment that may facilitate intellectual development, so that the genetic influence is mediated by a culturally determined environmental influence. In this case, it would be more proper to call the environmental mediating variable, rather than the genetic variable, causal.

Early versus late adversities

At various times in an individual's lifetime, they will encounter negative events or life stressors. A key question in developmental research is how the outcomes and coping strategies associated with these events differ as a function of when they occur. A young child may be particularly resilient to adversity; underdeveloped cognitive systems, for example, may restrict encoding of the event, such that its impact on future development is negligible. However, children are particularly vulnerable when the negative or traumatic event disrupts the normal course of development.

Over the last hundred years there have been several reports of children being rescued from severe adverse experiences, often involving considerable social and emotional neglect. Generally, the normal development of the child is severely compromised and, although improvements do occur with significant medical and psychiatric intervention, the child remains intellectually and socially impaired. These cases support the suggestion that children must attain specific developmental achievements within a critical age period.

These cases are, however, both extreme and rare. Many children experience some form of less severe adversity in childhood that may lead to behavioural disruption, but to what extent does this disruption persist – if at all? This will certainly depend on the nature of the negative event; a single short-lived adversity, such as a one-off hospitalization, while traumatic, may be less detrimental than sustained problems such as extreme poverty or an alcohol-dependent parent. However, individuals differ greatly in the extent to which they are vulnerable or resilient to the effects of negative events. Thus, it may be particularly important to consider why some individuals fare better than others following negative experience. Generally, events in early life will influence the development of risk factors such as self-concept, social skills and so on, while later events will act as causal factors given the existence of such risk or protective factors. This accounts for the apparent resilience found in some individuals where appropriate early experiences and subsequent emotional and cognitive development may act as protective factors.

Historical models and theories

Psychoanalytical theories

Psychoanalytical theories have common origins in the work of Sigmund Freud. They emphasize a role of unconscious as well as conscious processes in governing human behaviour. Early focus was on social and emotional characteristics of the adult and how these were shaped during early development. According to traditional theories, personality traits can be traced back to fixations held during various aspects of early development. An 'anal' personality type, for example, describes someone who is overly pedantic and compulsive, and would have had strong fixations in the anal stage of development. Parents and their methods of care are considered particularly important, with the infant playing a largely passive role in the outcome of their own development.

Social learning

A major component of social learning theory is learning through the observation and imitation of others, often termed 'observational learning' or 'modelling'. Although many imitated behaviours are reinforced in some way, studies by Albert Bandura in the 1960s and 1970s showed that children would mimic the behaviour of an observed model in the apparent absence of any positive or negative reinforcer. Imitation, however, does not occur spontaneously. Children appear to actively select the behaviours that they imitate. Contemporary social learning theorists are particularly interested in the factors that influence this selectivity, such as the personality or past experience of the child, or situational factors. More recently, social learning theory has been adapted to link closely with cognitive theories of development.

Jean Piaget's epistemological account of cognitive development is considered in more detail in other sections of this chapter. Based predominantly on the observations of his own children, Piaget developed a stage theory of cognitive development that emphasized the interaction between biological predispositions and environmental influence. His work also encompassed other aspects of development (e.g. moral development); however, it is his cognitive model of development that has been particularly influential. One of the most important contributions that Piaget made was to establish the child as playing a dynamic role in his or her own development, through interaction with physical and social worlds, that is driven by simple, inherited predispositions.

Summary

Understanding human development requires consideration of the relative influences of inherited biological dispositions and of environmental factors on phenotypic expression. This is referred to as the nature–nurture debate, and it is now accepted widely that genetic and environmental factors are not simply additive but interact in complex ways. Generally, it is believed that an individual may vary on any given characteristic within limits determined by their genotype, while the variation within these limits is determined by the environment.

Stage theories have been employed to explain many important aspects of human development. They describe qualitative changes that occur at various periods during development. Individuals pass through developmental stages in a fixed order, with attainment of the characteristic skills of one stage being required before the individual can pass on to the next. Although sometimes criticized for their rigidity, stage theories have been extremely influential and describe behaviours that are often readily observable in the developing child. A related concept is that of *maturational tasks*, which consolidate and expand current skills necessary for development on to higher levels of functioning.

It is important to be aware of several key theoretical frameworks, in particular psychoanalytical theory, social learning theory, and Piaget's stage theory of cognitive and intellectual development. These theories have been extremely influential in shaping understanding and research in human development.

METHODOLOGIES

Cross-sectional studies

Cross-sectional studies examine age-related differences in behaviour by comparing children of different age groups in the same situation and at the same point in time. This enables relatively quick and easy collection of large amounts of data. However, this method provides little information on how individual children develop or on the influ-

ence of past determinants on observed behaviour. It is possible to obtain self-reported retrospective accounts of past development, but this method is potentially flawed and has questionable reliability and validity. Furthermore, selection of different individuals means that comparison groups may differ on factors other than age. A 10-year-old child born in 1998, for example, may differ from a 5-year-old child born in 2003 due to different economic or social influences during their early development. Thus, there are limits on the conclusions that can be drawn from cross-sectional studies, and longitudinal or cohort methods are employed to tackle more complex questions such as individual stability over time.

Cohort studies

In cohort studies, individuals are chosen on the basis of one or more unifying characteristics. For example, studies examining the genetic and environmental risk factors for schizophrenia often enlist the children of schizophrenic patients, as the incidence of schizophrenia is elevated in these children compared with in the normal population. However, this presents problems relating to the generalizability of the findings. Longitudinal studies also require substantial resources to sustain them over the, often lengthy, study period. They also may achieve a less representative sample than a cross-sectional study, due to some people being unwilling to commit to repetitive testing over time. Furthermore, dropout rates are often high and such attrition may not be random.

Individual studies

On occasion it is beneficial to study an individual child in a detailed manner, which is not practical with larger groups of children. This method is particularly useful when studying unusual cases, such as rare diseases, where individual examples are uncommon. The fundamental limitation of such an idiographic approach, however, is an inability to generalize findings to the wider population; it is plausible that the observations are unique to the child being studied. Nevertheless, this method can provide useful insights that guide future research using more systematic techniques in a large number of individuals.

Identification and evaluation of influences

There is substantial interest in the relative influence of genetic and environmental factors on development. Methodologies for investigating this rely on comparison being made along certain phenotypic dimensions between individuals of varying genetic similarity. For example, identical (MZ) twins are genetically identical, while non-identical (DZ) twins share only, on average, one-half of their variegated genotype. As such, if a psychological characteristic is correlated more strongly in MZ twins compared with DZ twins, then this suggests a genetic influence. Several varying degrees of genetic and environmental similarity are possible: MZ twins reared together, MZ twins reared apart, DZ twins reared together, DZ twins reared apart, siblings reared together, siblings reared apart, cousins, and adopted (i.e. unrelated) children reared together. Other combinations are possible, but this demonstrates the variety of comparisons that may be made.

On a variety of phenotypic characteristics (e.g. personality and other psychological dimensions), the degree of similarity between individuals increases with genetic similarity (and also with environmental similarity). The interpretation of such results, however, is highly problematic, and it is becoming increasingly clear that accurate characterization and reliable measurement of the phenotypic characteristic of interest are of primary importance. Moreover, this basic methodology relies on a simple additive model of genetic and environmental influence that is probably incomplete, as it fails to take into account gene × environment interactions.

Summary

Three main methodologies have been employed to study human development, all of which have their own strengths and weaknesses. Cross-sectional studies are conducted at one point in time in groups of children of different ages. They provide a quick and easy method of collecting a large amount of data, but they do not enable deeper understanding of how individuals develop over time. Cohort studies recruit individuals based on a unifying characteristic; however, there are problems when generalizing these data to wider populations. Longitudinal studies involve the repeated testing of the same sample at various periods during their development, which allows analysis of individual development across time, but these studies are time-consuming and costly. Less commonly, a single child may be studied in depth, which is not possible with larger samples, although there are problems associated with the generalizability of the findings.

By studying individuals of varying genetic and environmental similarity, it is possible to examine the relative influence of these factors on phenotypic characteristics. However, interpretation of results from these studies can be problematic.

ATTACHMENT THEORY

Attachment theory attempts to describe the behaviour of infants (generally in primate species) and relate this to behaviour in adulthood, in particular focusing on the progression of these behaviours as they become focused on (usually) one primary caregiver, known as the attachment figure, through infancy, and the relationship between the resulting attachment style and interpersonal behaviour in adulthood. The main premise of attachment theory is that the

need for physical proximity and emotional closeness is an innate behaviour and a necessary precursor to and stimulus of psychological, emotional and social development. The primary attachment figure (usually, but not necessarily, the mother) serves as the model for the development of future relationships in later childhood and through adulthood.

Attachment behaviours comprise relevant behaviours directed towards the primary attachment figure, who is the focus of these behaviours (Table 8.1). In early life (typically the first 6 months), the infant does not display selectivity in the direction of relevant behaviours (e.g. seeking closeness), but after this early period these become increasingly directed towards an attachment figure. This represents the formation of an attachment with this individual; interestingly, it depends primarily not on obtaining resources (e.g. food) from that individual but instead on proximity, affection and so on. Therefore, direct survival behaviours such as feeding appear to be distinct from attachment behaviours such as seeking affection.

For example, rhesus monkeys separated from their mother will spend the majority of time clinging to a wool doll rather than a wire doll, even when it is the latter that provides food. This suggests that attachment behaviours are related not to feeding but to proximity, nurturance and so on cued by other stimuli (such as the physical characteristics of the mother).

It has been suggested widely that the initial attachment forms a template for the development of subsequent social and emotional relationships, so that early behaviour may predict later social behaviour.

Table 8.1 Development of attachment behaviours

Evolutionary basis	Infant's best hope for survival lies in proximity to caretaker
Early weeks	Infant predisposed to achieve closeness
3rd month	Attachment remains indiscriminate and transient
6th month	A single (usually) attachment figure is established
12th month	Fear of strangers, attachment maintained over distance
24th month	Attachment figure (usually mother) of primary importance
36th month	Substitute attachment figures are accepted in absence of primary
Adulthood	Early attachment behaviour related to later emotional development

Emotional development

Attachment behaviours in early life appear to be highly correlated with subsequent development and behaviour. Secure attachment styles are related to initiative-taking, social competence and the ability to form friendships in later childhood. Insecure attachment styles, on the other hand, are related to subsequent social withdrawal and difficulty in forming friendships. Attachment behaviour seems to be relatively stable, once established, over the lifespan.

Affect regulation

Insecure attachment behaviours include inability to express affection with the attachment figure, either by being excessively dependent and at the same time irritated by the attachment figure, or by displaying avoidant behaviour, and this may be related to affect regulation in adult life. Generally, insecure attachments have been found to be related to emotional disturbance in adolescence and adulthood, although the strength of the relationship is modest.

Relationships

The primary attachment figure serves as the model for subsequent relationships, including friendships and romantic relationships. Patterns of adult romantic relationships appear similar to attachment behaviours, with the proportion of adults describing themselves as within each category roughly corresponding to the proportions found in infant attachment styles. Retrospective questionnaires suggest that these styles of romantic affection are related to the perceived or remembered behaviour of the mother towards the individual.

Secure attachment

Infants who display secure attachment behaviour use the attachment figure as a base from which to explore, occasionally returning to seek affection. Separation from the attachment figure induces anxiety and distress and interaction is sought when reunited, after which anxiety reduces and exploration resumes. This is most evident in the first 36 months of life, after which separation anxiety reduces and attachment substitutes may be accepted.

Secure attachment behaviour is related to early infant temperament, child-focused caring, nurturing behaviour from the attachment figure, encouragement for the child to explore independently, and provision of a secure base by the attachment figure.

Insecure attachment

Insecure attachment behaviours comprise two distinct categories:

Insecure-avoidant behaviours are characterized by muted distress in the absence of the attachment figure and a

minimal response when reunited. What may appear to be secure attachment behaviour may in fact be insecure-avoidant behaviour, due to this apparent independence, and making the distinction may be difficult. The clearest difference lies in the fact that securely attached children still display closeness, balancing a desire for affection with independence, while insecure-avoidant children almost exclusively seek distance and self-sufficiency at the expense of closeness.

Insecure-ambivalent behaviours represent the least common attachment style and represent a mixture of excessively dependent behaviour towards the attachment figure with a lack of obvious affection. Separation results in excessive distress in the child but, when reunited, the child will then resist physical contact. Distress continues for some time after the return of the attachment figure.

There is evidence for a relationship between insecure attachment styles and future antisocial behaviour and affective disorders. If such attachment behaviours are identified in early life, then teaching parenting skills to the parents of the child, specifically to be more caring towards the child, may reduce this risk. However, it is simplistic to attribute attachment behaviours solely to the parenting style of the parents, and other factors such as early infant temperament appear relevant.

Early separation

Given the relationship between attachment behaviour and subsequent behaviour in adolescence and adulthood, separation from initial attachment figures is related to subsequent behavioural problems. This is related to the age at which such separation takes place. For example, selective attachments do not begin to form until 6 months of age, so that separation before this time may have relatively little effect, allowing the formation of attachment behaviours towards the adoptive caregiver. Similarly, after the first few years of life, the child will have developed a pattern of attachment behaviour, and therefore a model or template of relationships, so that separation will not result in excessive disturbance, as the child will be able to accept the adoptive caregiver as a substitute attachment figure.

The critical period appears to be between 6 and 36 months, when the greatest changes occur in the pattern of attachment behaviours. Separation during this period is likely to have the greatest impact on the child's subsequent behaviour, as it will disrupt the formation of normal attachment behaviour.

Failure to develop selective attachments

Autistic children are characterized by an unwillingness to seek physical contact with parents. This becomes most apparent between 12 and 36 months, when normal, selective attachments would otherwise typically form. This is highly correlated with similar behaviour in adult life and appears to be largely the result of physical and genetic factors rather than social or developmental processes.

The failure to develop selective attachments in autistic children as a result of internal factors should be distinguished from a breakdown of the normal attachment process because of environmental or rearing factors. In this case, the lack of a clear single caregiver results in incomplete development of attachment behaviours, so that strangers and parents are treated similarly, without any apparent stranger anxiety. At the same time, the behaviour towards such figures is superficial and any affection easily terminated, so that separation anxiety also does not occur. It is for this reason that early separation in the critical period is most detrimental to the child, in particular if the child is not placed in a caring adoptive or foster home.

Maternal bonding

The anxiety that results from separation of the child from the attachment figure is a necessary part of the development of secure attachments, since it is the eventual reunion that gives the child confidence in the attachment figure. This, in turn, gives the child confidence in relationships, understanding that they persist over time and distance. The term 'maternal bonding' is potentially inappropriate, however, in that, although usual in most Western cultures, it is not necessary that the primary attachment figure be the mother.

Summary

Attachment theory is a framework for understanding the behaviour of infant primates whereby proximity and affection are sought (attachment behaviours) and appear to act as primary drives, in a similar way to hunger and thirst. An attachment relationship develops between the infant and the care provider (attachment figure). This serves as the model for subsequent relationships of all kinds. Infant attachment behaviours appear to relate modestly to adolescent and adult friendships and romantic relationships. Selective attachments usually do not form before 6 months.

Attachment to an attachment figure may be either secure or insecure, with the latter further subdivided into avoidant and ambivalent. Secure attachment represents a relationship where the attachment figure provides a base for independent exploration, and, although separation from the attachment figure results in distress, this rapidly subsides on reunion. Insecure-avoidant attachment is characterized by a relative lack of distress on separation and a general tendency for affection to be muted. Apparent secureness masks an unwillingness to display closeness. Insecure-ambivalent behaviour is characterized by excessive dependence but an unwillingness to display affection towards the attachment figure. There is some evidence for a relationship between attachment style and subsequent antisocial behaviour and affective disturbance.

The development of attachment behaviours occurs between 6 and 36 months. This has implications for the separation of the child from the primary attachment figure (usually the mother), with this critical period being associated with the greatest impact on the child's behaviour if separation occurs.

A failure to develop selective factors may be the result of internal factors (e.g. autism) or external factors (e.g. disturbed family function). The latter situation may also be related to early separation of the child from the primary attachment figure during the critical period (i.e. the first 3 years of life).

FAMILY RELATIONSHIPS

Parental practices and attitudes

The behaviour of individuals towards their offspring is of primary importance in determining attachment behaviours in infants, with these factors further interacting with temperamental features of the infant. Parenting style is commonly characterized along two dimensions: restrictive-permissive and loving-hostile. Alternatively, three categories of parenting style have been suggested: permissive (warm and caring, while accepting unorthodox behaviour), authoritarian-restrictive (less emotionally close and highly controlling), and authoritative (enforce rules and demand achievement, yet warm and loving).

The development of the child depends greatly on the parenting style adopted by the child's parents. For example, authoritarian-restrictive parents tend to have highly submissive children, while authoritative parents tend to have independent children (largely as a result of the emphasis placed on expressiveness within set boundaries). The behaviour of the parents is most important in the early development of the child, and this establishes a pattern of behaviour in the child that is sustained over time. In particular, the *consistency* of the parents' behaviour towards the child, for example with reference to the rewards and punishments consequent on certain behaviour, is a strong predictor of subsequent emotional and social behaviour in the child. *Inconsistent* parenting behaviour is associated with negative outcomes along these dimensions, presumably because it inhibits the development of a clear social model to guide behaviour.

The extent to which specific attitudes of parents is mirrored in subsequent attitudes in their children is debatable: any relationship may be weak overall, and there are numerous examples of children adopting opposing attitudes and beliefs, in particular during adolescence.

Distorted family function

There are a variety of ways in which normal family function may be distorted. Such (dysfunctional) families are highly varied in the nature and extent of distorted function, so that care should be taken in making comparisons. Certainly there is no such thing as a 'typical' dysfunctional family. Distorted patterns of communication are often at the core of dysfunctional families; for example, positive or neutral remarks may be interpreted in a negative way, leading to potential discord. This situation may be self-sustaining, as individuals attempt to exert their influence in response to perceived insults by reacting in a similar way. Conversely, behaviour that would be useful in reducing discord (e.g. apology) is used rarely and is typically seen by those involved as an admission of defeat.

Overprotection may be the result of enmeshment within a family, in which case it is broadly applicable to all members of the family unit. It may also occur in specific cases; the central feature of overprotection is an unwillingness to allow a member of the family group to display individuality, especially outside of the family group. The characteristic feature of such family behaviour is an apparent excess of closeness that is selfishly motivated.

Rejection may be a specific individual effect, for example where a mother rejects closeness with her child, or a more general family effect. In this latter case, the family will not have a strong sense of collective identity, with individuals showing only weak attachment relationships to other members, often forming stronger attachments outside the family group. Isolation and loneliness commonly result.

Enmeshment refers to families where individual identity and individuality are lost to family status and role, in direct contrast to rejection. A high degree of homogeneity, especially of opinions and beliefs, results in little input in discussions regarding the family. Family structure is tightly defined, and any individual family member attempting to display independence is likely to have this suppressed.

Bereavement

Bereavement may involve the death of a loved one or relative, but more generally (and metaphorically) it can indicate loss of any kind (e.g. moving school, leaving home). This latter use is based on some evidence that the reaction to loss of any kind follows a similar pattern, although the intensity of these reactions will vary. Typical reactions to loss include: disbelief/denial, emotional blunting/numbness, and excessive rumination over the lost object or individual.

Considerable adjustment is required in any case of bereavement, and the effect on the child may be considerable. Behavioural problems are associated with the loss of a parent in children, and this is related to poor adjustment in the surviving parent, although the direction of causation may be complex. Subsequent development of the child may be influenced most by the subsequent behaviour of the surviving parent (e.g. change of parenting style, relative neglect due to increased time demands), rather than the loss of a parent specifically.

Divorce

The effects of divorce on child development are difficult to delineate, as parents who divorce have generally been in conflict for some time before the divorce. In general, in the first year following a divorce, the parents of the child become more permissive and less controlling, and communication between parents and child deteriorates. This behaviour occurs in both the parent without the child and the parent who cares for the child, but usually this change in parenting style resolves after the first year following the divorce. The effects of the divorce on the child depend on the age of the child, the degree of hostility resulting from the divorce, the use of the child by the conflicting parents to achieve their own aims, and the degree of adjustment of the parent who remains caring for the child.

In almost all cases there is a degree of distress in the child, allied to feelings of responsibility or guilt, and consequent depression or hostility. In most cases, however, this resolves relatively quickly (usually after the first year), provided that the parent with whom the child remains also adjusts over this period.

Intrafamilial abuse

Abuse of children is most dangerous in the early years of life (under 4 years), not least because this is when the child is physically most vulnerable. This is also the period when the antecedents of self-esteem and self-image are developing. Immediate effects of intrafamilial abuse (either physical or sexual) include sleep disturbance, eating disturbance, depression, phobias or anxiety, behavioural problems and social problems. Longer-term effects are more difficult to delineate, but there is some evidence that initial consequences may persist for several years if abuse continues. In particular, several problems appear to persist, including depression (especially in sexual abuse), self-harm, low self-esteem and impaired relationship formation.

These effects are greatest if the abuse comes from within the family group (which is the most common source), if the abuse is sustained over a substantial period, and if the abused is not believed when help is sought.

Non-orthodox family structures

The most common non-orthodox family structure is the single-parent family, a situation that has increased in prevalence recently. This may be the result of bereavement (uncommon), divorce (common) or single motherhood (also common). The consequences of living with only a mother (i.e. single-mother families) have been investigated extensively, since this is the most typical single-parent situation. Effects have been reported for self-identity (boys, in particular, may have problems achieving a clear sexual identity if the father is absent in early life), gratification (boys, again, are most affected and appear to show impaired ability to delay gratification and control impulses), and social skills (interaction with the opposite sex, in particular in girls, appears to be impaired).

There is also evidence that the relative difficulty in forming relationships with the opposite sex found in the children of single mothers persists into adulthood, although the effect is weak. The potential confound of differences in parenting behaviour existing in single-parent families is difficult to control for in studies that consider the impact of single-parent families on children, and it is difficult to conclude that behavioural and psychological correlates of a single parent upbringing are a consequence of the fact that only one parent was present. Similar consequences for child development have been found in other, rare family structures (e.g. children raised by grandparents).

Although the extended family structure found in some cultures and ethnic groups should not be regarded as unorthodox, it is not the common structure in Western cultures. In general, children raised in such families tend to develop high levels of self-esteem and the family structure seems to represent a protective factor (i.e. by providing high levels of social support).

Summary

Parenting style is an important determinant of child behaviour. The behaviour of the parents may be described as either dimensional (e.g. restrictive-permissive/loving-hostile) or categorical (e.g. permissive, restrictive, authoritative). Parenting style correlates with the behaviour of the child. For example, restrictive parents tend to have submissive children. This relationship is statistical, so that it is not possible to accurately predict behaviour in the child from the behaviour of the parents in specific cases.

The impact of bereavement and divorce on family function, in particular children's behaviour, may be regarded under the more general term of loss. There are certain similarities between bereavement and divorce, although there are also unique characteristics of each. Distress following loss of any kind is normal and only problematic if it becomes chronic in nature and extends beyond the normal time for adaptation and coping to take place.

Specific terms are used to describe different features of distorted family function, such as overprotection, rejection and enmeshment. Although there is no such thing as a typical dysfunctional family, and the causes of dysfunction may vary widely, certain characteristics appear repeatedly (e.g. distorted or impoverished communication between family members).

TEMPERAMENT

Temperament and parent–child relationships

From birth, infants display apparently innate differences in behaviour. They differ in their moods, activity levels,

stability of sleeping and feeding patterns, sociability, and response to novel situations. These early biases are known as differences in temperament. Given that these differences are observed soon after birth, it is reasonable to suggest that differences in temperament are, at least in part, determined genetically.

Temperament also influences the relationship that forms between the child and the parent. A parent of a 'difficult' child may display irritability and frustration, which, in turn, increases the irritability of the child. In extreme cases, the stress of dealing with a difficult child may interfere with the natural attachment process, particularly in families where there is limited external support. Conversely, a parent is more likely to engage with a child who is sociable and quick to smile. Consequently, even within the same household, parents may encourage different patterns of development in their children, driven by differences in the children's temperament. Parenting style itself, however, is subject to change. Parents of a difficult child may learn to adapt and develop new and more fruitful ways of interacting with their offspring, so that even when bonding is initially compromised this can be rectified with appropriate parental adjustment.

Origins, typologies and stability of temperament

Early work on temperament identified several dimensions along which young children vary. The analysis of these dimensions resulted in three typologies of temperament: *easy children* (about 40%), characterized as having regular sleeping and feeding patterns, sociability, enjoyment of physical contact, and adaptability to new situations; *difficult children* (about 10%), characterized by irregular sleeping and feeding patterns, irritability, emotional lability and negative responding to new situations; and *slow-to-warm-up children* (about 15%), characterized by relatively low activity, slowness to adapt to new situations, and treating new situations with initial suspicion. A proportion cannot be reliably classified into one of these groups.

Around one-third of the original sample of children studied in this research did not clearly fall into any of these groups, but follow-up work suggested some degree of stability in temperament. Children identified as 'difficult' may not always be classified this way but are more likely to have problems in later life.

Temperament theory suggests that the nature and behaviour of a child is not simply a product of his or her environment (e.g. parenting style). Instead, the child interacts with the world and, in doing so, modifies behaviour directed towards them. The goodness-of-fit of the child's environment should be taken into account – a 'difficult' child may prompt negative responses in his or her parents. Alternatively, parents may work hard to overcome the challenges of dealing with a 'difficult' child, handling the child with emphasized patience and understanding. The latter case is more likely to result in healthy development of the child.

Temperament and personality

A critical question in developmental psychology is the extent to which personality in adulthood may be predicted by infant temperament. Although research has suggested links between early temperament and personality in later life, any continuity that does exist appears to be weak, presumably because of the strong environmental and societal influences on the development of personality. However, temperament does play an important role in the emerging personality of the individual; it sets a framework of behaviour, but personality will be a result of how this basic pattern of behaviour is affected by maturation and experience.

Temperament × environment interactions may serve to strengthen innate dispositions and mould personality. A sociable child, for example, will seek contact with others, in turn promoting sociability and social development. Furthermore, temperament is associated with various long-term outcomes, including the quality of familial and peer relationships, psychopathology and psychological adjustment.

There is some evidence to suggest that temperament of very young children is related to subsequent attachment behaviour (over a relatively short timespan): *easy* with secure attachment, *difficult* with anxious-avoidant attachment, and *slow-to-warm* up with anxious-ambivalent attachment. Attachment behaviour, in turn, has been shown to be weakly associated with adult behaviour. The greater the temporal distance between the temperament and current behaviour, the weaker the relationship.

Vulnerability and resilience

Psychosocial adversity and negative life events are associated with higher incidence of psychiatric disorder. However, individuals differ in the extent to which they are vulnerable or resilient to long-term negative impact following adverse experiences. Several factors have been identified that may increase vulnerability, such as family discord, intrafamilial abuse and early bereavement (particularly of the primary attachment figure).

Some children show great resilience to life stressors. Those that demonstrate secure attachment behaviour, for example, are likely to be able to form replacement attachments if the mother is lost (depending on the age at which the loss occurs). It is also possible that experience of early stress may result in resilience to stress in later life in some individuals but an increase in vulnerability in others.

The long-term impact of traumatic events will depend on the temperamental disposition of the child. Difficult and slow-to-warm-up children are more vulnerable to stressors and appear to be more at risk of subsequent deviant

behaviours, including psychiatric problems such as depression, hysteria, neuroticism and psychosis.

Summary

From birth, infants display differences in behaviour that are referred to as individual *temperaments*. Children with 'difficult' temperaments are more irritable, less content and less sociable than children with 'easy' temperaments. An intermediate temperament, 'slow-to-warm-up', has been identified that describes children who may show initially negative emotional reactions that improve with time and experience. Development of temperament may be related to the development of attachment style in early infancy.

Some psychiatric disorders (e.g. unipolar depression) may develop as a result of certain environmental risk factors. However, individuals differ in the extent to which the presence of these risk factors is likely to result in the expression of psychiatric disturbance: some people seem particularly vulnerable, while others are resilient to these effects. The mechanisms underlying these differences are unclear but are likely to include the coping style of the individual.

COGNITIVE DEVELOPMENT

Jean Piaget developed an epistemological account of intellectual development that transformed our understanding of the way in which children think. He argued that the general pattern of cognitive growth was universal, and interaction with objects and people in the world forced children to follow comparable patterns of development. Piaget was particularly interested in *how* children think, the processes that underlie cognition, and how these processes change with age. An important feature of his theory, and one that is still largely accepted today, is that children's thought processes are qualitatively, not just quantitatively, different from adults'.

According to this model, children are not merely passive observers of the world. Cognitive development progresses via the interactive processes of *assimilation* and *accommodation*. At birth a child inherits relatively primitive mental structures, largely reflexive in nature, which enable basic interaction with the environment and which are modified with experience. First, the child interacts with the environment and interprets information in the context of existing mental structures (assimilation). Then, when presented with new information or objects that do not fit with existing cognitive schema, the child develops his or her repertoire of actions based on the environment (accommodation). Thus, in Piaget's view, cognitive development comprised a continuous interaction between the child and the environment, in which knowledge is repeatedly constructed and reconstructed in light of new experience.

Piaget argued that cognitive development was not a gradual process but advanced in a step-like fashion. He described specific, qualitatively distinct stages of development in which accomplishment of critical aspects of each stage was required before the child could advance to subsequent stages. These stages described a global process of cognitive development comprising the totality of cognitive growth from birth to maturity and are summarized in Table 8.2.

These transitions between stages suggest that, at critical periods, the child undergoes a comprehensive mental reorganization that encompasses all aspects of cognition. The ages described provide a guide to when these transitions occur, but children may vary considerably in the ages that they progress from one stage to another, with some individuals failing to achieve the final stage entirely.

Some key concepts should be noted that further characterize the different stages in development. *Object permanence* describes the acquired ability of the child to understand that objects in the environment exist even when

Table 8.2 Stages in Piaget's Model of Cognitive Epistemology

Sensorimotor (birth–18 months)	Knowledge of the world is acquired primarily through sensory experience and basic motor actions (sucking, grasping, biting), which become more complex and coordinated over time. Only towards the end of this stage does the child begin to form internal mental representations of the world and begin to engage in intentional behaviour
Preoperational (18 months–7 years)	Characterized by the child beginning to engage in symbolic thought (imaginative play, drawing), where words/mental images stand for objects and people. Focus on static states rather than transformations
Concrete operations (7–11 years)	Logical thought and problem-solving ability emerges. Children acquire a number of cognitive operations that were not possible at earlier stages, including ability to see situations from the viewpoint of others, reversible thinking and ability to understand transformations. Abstract conceptualizations, however, are limited
Formal operational (11 years+)	Children can think in terms of abstract and hypothetical concepts. Logical and rational thought develop so that the child can conceive possible outcomes of actions and perform advanced problem-solving by constructing and testing hypotheses systematically. Complex self-identity develops

he or she can no longer see them. This ability is one of the most important developments during the sensorimotor stage. Preoperational children have a well-developed sense of object permanence but are largely *egocentric* – that is, they can only construe the world from their own viewpoint and lack understanding that alternative perspectives exist. Egocentrism is lost by the concrete operational stage, in which the most important acquired skill is *conservation*. This refers to an understanding that the fundamental nature of an object can remain constant despite changes in appearance. The classic example is that children in this stage understand that a set volume of liquid does not change when poured into different-shaped containers.

A particularly common criticism of Piaget's model is that he was overly pessimistic when describing the cognitive abilities of children. He believed that biological factors constrained the order and rate at which a child was able to develop. However, abilities such as object permanence have been observed in children much younger than Piaget would have predicted. Underdeveloped motor ability or difficulty in relaying task instructions to young children may have contributed to these underestimations in cognitive capability. In addition, the stage-like nature of development proposed by Piaget has been challenged. It appears that development may be more gradual and continuous than the model suggested. For example, children may show some aspects of object permanence or non-egocentric behaviour in some circumstances but not others, arguing against the abrupt changes in development associated with a stage theory.

Despite these limitations, Piaget's theory has been hugely influential due to its depth, testable hypotheses, and the observational and empirical basis of much of the theory. It provides a framework of global cognitive and intellectual development that explains the qualitative differences in thinking during development and also *how* changes occur. The assumption that children are not passive observers of the world has also had important implications for education.

Thought and communication

Piaget claimed that the acquisition of language progresses as part of the natural development of cognition, and therefore the way in which a child communicates with others depends on the developmental stage the child is at.

The first steps on the road to acquiring language occur with the emergence of symbolic thought towards the end of the sensorimotor stage. However, it is during the preoperational period that the child begins to display significant advances in linguistic skills, from single-word utterances, through short sentences, to competent, if unsophisticated, language. Towards the end of the preoperational stage, the child begins to learn and apply the *rules* of language (*language pragmatics*) as a means of more sophisticated communication. At this stage, children begin to show adult-like gestures associated with verbal communication and adapt their communication to the situation they are in or the person to whom they are talking. This social skill improves with the reduction of egocentrism, when children can appreciate the perspective of others and thus adapt communication appropriately. A noticeable change that occurs during the preoperational stage is the development of politeness in the linguistic competence of these children. More complex and subtle features of communication, such as indirect questions, hints and sarcasm, require a certain level of cognitive maturity before they can be learned.

The loss of egocentrism during the *concrete operational* stage characterizes a dynamic change in communication. Understanding the perspective of others means that, for the first time, the child appreciates that another person may not be interested in some topics of conversation and can identify what another person does or does not know. Persuasive communication marks another important development in linguistic ability; during earlier development, children tend to repeat crude requests, while concrete operational children form persuasive arguments that take into account the attitudes or desires of the person to whom the request is directed. During this stage, humour also develops, with children beginning to understand and tell rudimentary jokes.

Finally, the *formal operations* stage, in the context of communication, represents the final development of the social skills that must be allied to linguistic ability. An increasingly level of subtlety is evident in requests and persuasion, and sophistication in the use of humour (sarcasm, irony) and pragmatics appears. Children also acquire the ability to infer meaning of unfamiliar words by the context in which the word appears, without the need for explicit definition.

Therefore, the linguistic ability and communicative competence of children vary as a function of their level of cognitive development. As the capacity of abstraction develops, egocentrism declines and communicative social skills (pragmatics) develop, spoken language becomes more complex and goal-oriented, particularly in social situations.

Summary

Piaget's model of cognitive development proposes that children pass through step-like stages of cognitive development that are qualitatively different and proceed in an invariant order. These stages are the *sensorimotor*, *preoperational*, *concrete operational* and *formal operational*. To understand the model fully, one should be aware of the characteristics that define each stage and the approximate ages at which children move between them.

Although no longer accepted universally, Piaget's model remains influential. Key concepts include *assimilation*, *accommodation*, *operations* and *egocentricism*. These represent key mechanisms that allow the child to complete one stage of development and proceed on to the next.