## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreword</td>
<td>vii</td>
</tr>
<tr>
<td>Authors</td>
<td>viii</td>
</tr>
<tr>
<td>Contributors</td>
<td>ix</td>
</tr>
<tr>
<td>Reviewers</td>
<td>xiii</td>
</tr>
<tr>
<td>Preface</td>
<td>xiv</td>
</tr>
<tr>
<td>1 Introduction to evidence-based practice</td>
<td>1</td>
</tr>
<tr>
<td>Tammy Hoffmann, Sally Bennett and Chris Del Mar</td>
<td></td>
</tr>
<tr>
<td>2 Information needs, asking questions, and some basics of research studies</td>
<td>16</td>
</tr>
<tr>
<td>Chris Del Mar, Tammy Hoffmann and Paul Glasziou</td>
<td></td>
</tr>
<tr>
<td>3 Finding the evidence</td>
<td>41</td>
</tr>
<tr>
<td>Andrew Booth</td>
<td></td>
</tr>
<tr>
<td>4 Evidence about effects of interventions</td>
<td>64</td>
</tr>
<tr>
<td>Sally Bennett and Tammy Hoffmann</td>
<td></td>
</tr>
<tr>
<td>5 Questions about the effects of interventions: examples of appraisals from different health professions</td>
<td>102</td>
</tr>
<tr>
<td>Tammy Hoffmann, John Bennett, Jeff Coombes, Ronald Castelino, Edzard Ernst, Isabelle Jalbert, Karl Landorf, Michelle Miller, Shannon Munteanu, Nina Rødland Olsen, Emma Power, Gavin Smith, Alison Twycross and Caroline Wright</td>
<td></td>
</tr>
<tr>
<td>6 Evidence about diagnosis</td>
<td>136</td>
</tr>
<tr>
<td>Jenny Doust and Sharon Sanders</td>
<td></td>
</tr>
<tr>
<td>7 Questions about diagnosis: examples of appraisals from different health professions</td>
<td>159</td>
</tr>
<tr>
<td>Chris Del Mar, Sally Bennett, Bridget Abell, Malcolm Boyle, Kylie Hill, Samantha Hollingworth, Courtney Lewis, Gustavo Noffs, Claire Rickard, Sharon Sanders, Michal Schneider, Inge Stegeman and Adam Vogel</td>
<td></td>
</tr>
<tr>
<td>8 Evidence about prognosis</td>
<td>185</td>
</tr>
<tr>
<td>Julie K Tilson</td>
<td></td>
</tr>
<tr>
<td>9 Questions about prognosis: examples of appraisals from different health professions</td>
<td>202</td>
</tr>
<tr>
<td>Tammy Hoffmann, Chris Del Mar, Marilyn Baird, Malcolm Boyle, Jeff Coombes, Elizabeth Gibson, Karl Landorf, Shannon Munteanu, Lisa Nissen, Claire Rickard, Sheena Reilly, Sharon Sanders, Nabita Singh and Julie Tilson</td>
<td></td>
</tr>
<tr>
<td>10 Understanding evidence from qualitative research</td>
<td>226</td>
</tr>
<tr>
<td>Karin Hannes and Sally Bennett</td>
<td></td>
</tr>
<tr>
<td>11 Understanding evidence from qualitative research: examples of assessment of quality (critical appraisal) from different health professions</td>
<td>248</td>
</tr>
<tr>
<td>Sally Bennett, Bridget Abell, John Bennett, Fiona Bogossian, Carolyn Ee, Paul Jennings, Lisa Nissen, Rebecca Nund and Cylie Williams</td>
<td></td>
</tr>
<tr>
<td>12 Appraising and interpreting systematic reviews</td>
<td>292</td>
</tr>
<tr>
<td>Sally Bennett, Karin Hannes and Denise O’Connor</td>
<td></td>
</tr>
<tr>
<td>13 Clinical practice guidelines</td>
<td>323</td>
</tr>
<tr>
<td>Tammy Hoffmann and Heather Buchan</td>
<td></td>
</tr>
<tr>
<td>14 Shared decision making</td>
<td>337</td>
</tr>
<tr>
<td>Tammy Hoffmann and Leigh Tooth</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Clinical reasoning and evidence-based practice</td>
</tr>
<tr>
<td>----</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Merrill Turpin and Joy Higgs</td>
</tr>
<tr>
<td>16</td>
<td>Implementing evidence</td>
</tr>
<tr>
<td></td>
<td>Annie McCluskey and Denise O’Connor</td>
</tr>
<tr>
<td>17</td>
<td>Embedding evidence-based practice into routine clinical care</td>
</tr>
<tr>
<td></td>
<td>Ian Scott, Chris Del Mar, Tammy Hoffmann and Sally Bennett</td>
</tr>
<tr>
<td></td>
<td>Index</td>
</tr>
</tbody>
</table>
Evidence-based practice has been the subject of a new wave of criticism in the last few years. One of the criticisms, for example, is that it reduces the complexity of clinical practice to rule-based practice. But the original aim of evidence-based practice was much more modest – simply a hope that people would think about how they were thinking when it came to making healthcare decisions.

This intention remains. And more so than ever, both clinicians and their patients are encouraged to reflect on their thinking about health decisions – and, for example, consider what benefits and harms one proposed clinical management option has over an alternative. In evidence-based practice, there has been increasing interest in and efforts to incorporate patients’ values and circumstances into the decision-making process. This, of course, cannot be ‘rule-based’; this new edition clearly incorporates this message throughout, highlighting that the evidence always needs to be interpreted in context.

Evidence-based decision making remains relevant and common to all health professions and provides an ideal platform for interdisciplinary work. Different professions bring different perspectives to bear on the evidence, on relating the evidence to the individual, and on helping the individual reflect on their values as they consider the evidence for the options presented to them. This book provides examples of the use of evidence from a wide range of professions. The importance of this lies in making the practice of evidence-based practice clearly relevant to and achievable by all health professionals, whatever their unique role is in providing health care.

The editors have once again superbly cleared away the mystique of the term ‘evidence-based practice’ and made the concepts and processes clear and accessible. Throughout this book, the focus is on what it is that clinicians and patients really need to know to make evidence-informed decisions. Evidence-Based Practice across the Health Professions should help us all to incorporate the best available evidence into clinical and health service decisions.

Professor Sir Muir Gray, CBE MD
Director, Better Value Healthcare
Formerly Chief Knowledge Officer of the National Health Service (NHS), UK
Tammy Hoffmann  BOccThy (Hons 1), PhD
Professor of Clinical Epidemiology, Centre for Research in Evidence-Based Practice, Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Queensland, Australia
Tammy has been teaching and researching about evidence-based practice and shared decision making for over 15 years. She has an international reputation in her various areas of research. Her research spans many aspects of shared decision making, evidence-based practice, informed health decisions, improving the reporting and uptake of effective interventions, knowledge translation, minimising waste in research, and the teaching of evidence-based practice.

Sally Bennett  BOccThy (Hons), PhD
Associate Professor in Occupational Therapy, School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, Queensland, Australia
Sally has extensive experience in teaching and researching about evidence-based practice and knowledge translation. Her research interests include building capacity for knowledge translation, and translating knowledge for care of people with dementia. She was one of the leaders of the internationally recognised OTseeker database, which provides evidence relevant to occupational therapy. Sally is actively involved at the professional level, being on the editorial board of the Australian Occupational Therapy Journal and The Canadian Journal of Occupational Therapy.

Chris Del Mar  BSc, MA, MB BChir, MD, FRACGP, FAFPHM
Professor of Public Health and academic General Practitioner, Centre for Research in Evidence-Based Practice, Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Queensland, Australia
Chris worked as a full-time GP for many years before becoming Professor of General Practice at the University of Queensland. He was invited to become the Dean of a new Health Sciences and Medicine Faculty to start a new medical program and also became Pro-Vice Chancellor (Research) at Bond University. After overseeing the graduation of the first cohort of medical students, he stepped back from those roles to return to research and teaching. He has an international reputation in the management of acute respiratory infections (and is coordinating editor of the Acute Respiratory Infections Cochrane Group); general practice research; evidence-based medicine and systematic reviews; and randomised controlled trials, in both clinical medicine and health services research.
CONTRIBUTORS

Bridget Abell  BAppSc (HMS-Exercise Science), MSc
PhD Candidate, Centre for Research in Evidence-Based Practice, Bond University, Gold Coast, Queensland, Australia

Marilyn Baird  BA, DCR, PhD
Professor and Head, Department of Medical Imaging and Radiation Sciences, and Associate Dean, Learning and Teaching (Faculty of Medicine, Nursing and Health Sciences), Monash University, Melbourne, Victoria, Australia

John Bennett  BMedSc, MBBS, BA (Hons), PhD, FRACGP, FACHI
Discipline of General Practice, School of Medicine, and The University of Queensland Health Service, Brisbane, Queensland, Australia

Fiona Bogossian  RN, RM, DipAppSci (NEd), BAAppSci, MPH, PhD, FACM
Associate Professor, School of Nursing, Midwifery and Social Work, The University of Queensland, Brisbane, Queensland, Australia

Andrew Booth  BA, Dip Lib, MSc, MCLIP, PhD
Reader in Evidence Based Information Practice and Director of Information, Information Resources Group, HEDS, ScHARR, The University of Sheffield, Sheffield, United Kingdom

Malcolm Boyle  ADipBus, ADipHSc (Amb Off), MICA Cert, BInfoTech, MClinEpi, PhD
Associate Professor and Academic Lead in Paramedicine Education, School of Medicine, Griffith University, Gold Coast, Queensland, Australia

Heather Buchan  MBChB, MSc; FAFPHM
Senior Medical Advisor, Australian Commission on Safety and Quality in Health Care, Sydney, New South Wales, Australia

Ronald Castelino  BPharm, MPharm, PhD
Senior Lecturer in Pharmacology & Clinical Pharmacy, The University of Sydney, Sydney, New South Wales, Australia

Jeff Coombes  BEd (Hons), BAppSc, MEd, PhD
Professor in Exercise Science, School of Human Movement and Nutrition Sciences, The University of Queensland, Brisbane, Queensland, Australia

Jenny Doust  BEcons, BMBS, GradDipClinEpi, PhD, FRACGP
Professor of Clinical Epidemiology, Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Queensland, Australia
Carolyn Ee MBBS, PhD, MBBS, FRACGP, BAppSci (Chinese Med), MMed, Grad Cert Med Acup, PhD
Senior Research Fellow, National Institute of Complementary Medicine, Western Sydney University, Campbelltown, New South Wales, Australia

Edzard Ernst MD, PhD, FMedSci, FRCP, FRCP (Ed)
Emeritus Professor of Complementary Medicine, University of Exeter, Exeter, United Kingdom

Elizabeth Gibson BOccThy, PhD
Senior Research Fellow, Centre for Research in Evidence-Based Practice, Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Queensland, Australia

Paul Glasziou MBBS, PhD, FRACGP, MRCGP
Professor of Evidence-Based Medicine and Director of the Centre for Research in Evidence-Based Practice, Bond University, Gold Coast, Queensland, Australia

Karin Hannes BEdu, MSc Edu, MSc Med, PhD
Associate Professor in Qualitative Methodology and Meta-Synthesis at the Social Research Methodology Group, Centre for Sociological Research, Faculty of Social Sciences, University of Leuven, Leuven, Belgium

Joy Higgs AM, BSc, MHPEd, PhD, PFHEA
Professor in Higher Education, Charles Sturt University, Sydney, New South Wales, Australia

Kylie Hill BSc (Physiotherapy), PhD
Associate Professor, School of Physiotherapy and Exercise Science, Faculty of Health Sciences, Curtin University, Perth, Western Australia

Samantha Hollingworth BSc (Hons 1), PhD, MPH
Senior Lecturer, School of Pharmacy, The University of Queensland, Brisbane, Queensland, Australia

Isabelle Jalbert OD, MPH, PhD
Senior Lecturer, Deputy Head of School, School of Optometry and Vision Science, University of New South Wales, Sydney, New South Wales, Australia

Paul A Jennings BNur, GradCertBiostats, GCHPE, MClinEpi, PhD, FPA
Adjunct Professor, College of Health and Biomedicine, Victoria University, Melbourne, Victoria, Australia

Karl Landorf DipAppSc, GradCertClinInstr, GradDipEd, PhD
Associate Professor and Research Coordinator, School of Allied Health and La Trobe Sport and Exercise Medicine Research Centre, La Trobe University, Melbourne, Victoria, Australia

Courtney Lewis BSc, Msc
Speech pathologist and PhD candidate, Centre for Neuroscience of Speech, The University of Melbourne, Melbourne, Victoria, Australia
Annie McCluskey DipCOT, MA, PhD
Honorary Senior Lecturer, Faculty of Health Sciences, The University of Sydney, Sydney, New South Wales, Australia

Michelle Miller BSc, MNutDiet, PhD, AdvAPD
Professor and Head, Nutrition & Dietetics, Flinders University, Adelaide, South Australia, Australia

Shannon Munteanu BPod (Hons), PhD
Discipline Leader – Podiatry, Senior Lecturer, School of Allied Health and La Trobe Sport and Exercise Medicine Research Centre, La Trobe University, Melbourne, Victoria, Australia

Lisa Nissen BPharm, PhD, AdvPracPharm, FPS, FHKaph, FSHP
Professor (Head), School of Clinical Sciences, Queensland University of Technology, Brisbane, Queensland, Australia

Gustavo Noffs MSc, MD
ENT surgeon and PhD Candidate, Centre for Neuroscience of Speech, The University of Melbourne, Melbourne, Victoria, Australia

Rebecca Nund BSpPath (Hons), GCHEd, PhD
Lecturer in Speech Pathology, School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, Queensland, Australia

Denise O’Connor BAppScOT (Hons), PhD
Senior Research Fellow, School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia

Emma Power BAppSc (Speech Path, Hons 1), PhD
Senior Lecturer, Speech Pathology, Faculty of Health Sciences, University of Sydney, Sydney, New South Wales, Australia

Sheena Reilly FASSA, FAHMS, FRCSLT, FSPA, PhD
Professor and Pro Vice Chancellor (Health), Griffith University, Gold Coast, Queensland, Australia

Claire Rickard BN, RN, GradDip (CriticalCare), PhD, FAHMS, FACN
Professor and Director of Alliance for Vascular Access Teaching & Research Group, School of Nursing and Midwifery, Griffith University, Brisbane, Queensland, Australia

Nina Rydland Olsen PT, PhD
Associate Professor, Department of Occupational Therapy, Physiotherapy and Radiography, Faculty of Health and Social Sciences, Western Norway University of Applied Sciences, Bergen, Norway

Sharon Sanders BSc (Pod), MPH, PhD
Assistant Professor, Centre for Research in Evidence-Based Practice, Bond University, Gold Coast, Queensland, Australia
Michal Schneider  BSc, DipEd, MRepSc, GradCertHealthProfEdu, PhD
Deputy Head, Director of Research, Department of Medical Imaging and Radiation Sciences, Monash University, Melbourne, Victoria, Australia

Ian Scott  FRACP, MHA, MEd
Director of Internal Medicine and Clinical Epidemiology, Princess Alexandra Hospital, Brisbane, and Associate Professor of Medicine, School of Clinical Medicine, The University of Queensland, Brisbane, Queensland, Australia

Nabita Singh,  BSc (Biomedical Science), MBioEnt (Hons)
Assistant Lecturer, Department of Medical Imaging and Radiation Sciences, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Victoria, Australia

Gavin Smith  B Paramedic Studies, GradDip (EmergHealth), MEH, PhD, FPA
Associate Professor (Paramedic Science), College of Health and Biomedicine, Victoria University, Melbourne, Victoria, Australia

Inge Stegeman  BOptom, MSc Prevention and Public Health, PhD
Epidemiologist, Assistant Professor, University Medical Center, Utrecht, and Department of Otorhinolaryngology – Head and Neck Surgery, Julius Center for Health Sciences and Primary Care, Cochrane Netherlands University Medical Centre, Utrecht, The Netherlands

Julie K. Tilson,  PT, DPT, MS, NCS
Associate Professor of Clinical Physical Therapy, University of Southern California, Los Angeles, California, United States

Leigh Tooth  BOccThy (Hons), PhD
Associate Professor, Principal Research Fellow and Deputy Director of the Australian Longitudinal Study on Women’s Health, School of Public Health, The University of Queensland, Brisbane, Queensland, Australia

Merrill Turpin  BOccThy, GradDipCounsel, PhD
Senior Lecturer, School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, Queensland, Australia

Alison Twycross  PhD, MSc, DMS, CertEd (HE), RGN, RMN, RSCN
Head of Department and Professor of Children’s Nursing, London South Bank University, London, United Kingdom

Adam P Vogel  BA, MSc (SpPth), PhD
NHMRC Career Development Fellow and Director, Centre for Neuroscience of Speech, The University of Melbourne, Melbourne, Victoria, Australia; Hertie Institute for Clinical Brain Research, University of Tübingen, Germany

Cylie Williams  BAppSc (Pod), MHlthSc (Hlth Prom & Ed), PhD
Senior Research Fellow, Peninsula Health and Monash University, Melbourne, Victoria, Australia

Caroline Wright  DCR (T), BSc (Hons), MSc, PGCE, PhD
Senior Lecturer and Course Convenor Master of Radiation Therapy, School of Primary and Applied Health Care, Monash University, Melbourne, Victoria, Australia
REVIEWERS

Ronald L Castelino  BPharm, MPharm, PhD
Senior Lecturer, The University of Sydney, Sydney, New South Wales, Australia

Caleb Ferguson RN, BScN, MHlth, PhD, MACN
Chancellor’s Postdoctoral Research Fellow, Centre for Cardiovascular and Chronic Care, Faculty of Health, University of Technology Sydney, Sydney, New South Wales, Australia

Steven A Frost BNurs, ICU Cert, MPH, PhD
Lecturer, Intensive Care, Liverpool Hospital, and Western Sydney University, Campbelltown, Sydney, New South Wales, Australia

Rosemary King BA, MA, RN, RM
PhD Candidate. Lecturer in Midwifery, School of Nursing, Midwifery and Healthcare, Federation University, Mt Helen, Victoria, Australia

Irene Koren BA (Hons), BScN, MSc
Assistant Professor, School of Nursing, Laurentian University, Sudbury, Ontario, Canada

David Lim DrPH
Rural Interprofessional Education Team Leader, Flinders Rural Health SA; Faculty of Medicine, Nursing and Health Sciences, Flinders University, Adelaide, South Australia, Australia

Steve Milanese BAppSc (Physio), GradCert (SportsPhysio), GradDip (Ergon), MAppSc (ManipPhysio), PhD
Senior Lecturer in Physiotherapy, Program Director: Health Sciences Honours Program, Research Lead: Musculoskeletal Therapy, International Centre for Allied Health Evidence, Sansom Centre, University of South Australia, Adelaide, South Australia, Australia

Heleen Reid MHSc (OccTher)
Program Leader – Occupational Science and Therapy, School of Clinical Sciences, Auckland University of Technology, Auckland, New Zealand
Evidence-based practice, like the evidence itself, evolves with time. This led to the need to update the content and produce a third edition of this book so that it remains a contemporary guide to modern evidence-based practice.

Each time that we prepare a new edition, it is a great opportunity to incorporate the key developments in the field that have occurred in the few short years between editions. As evidence-based practice celebrates its 25-year anniversary, many changes have occurred over that time. In the first edition of this book, we very deliberately included a chapter about communicating evidence to and sharing decisions with patients. We did this because we passionately believe that this is a crucial step of evidence-based practice. At the time, it was quite unheard of for this topic to feature as part of a book about evidence-based practice. Gratifyingly, since we prepared the first edition, there has been increasing recognition and acceptance of the need to discuss evidence with patients, consider patient values, and collaborate with them during decision making as part of evidence-based practice.

That is just one example of how the field of evidence-based practice has matured. Certainly, many challenges to further improve evidence-based practice remain, but there are numerous talented, innovative and dedicated academics around the world who are actively working on ways to improve the very important contribution that evidence-based practice makes to health care. Many of the contributors to this edition are international and national leaders of evidence-based practice in their disciplines. We are indebted to them for their contributions to this book and for helping us to produce a truly inter-disciplinary book with 15 disciplines represented.

We hope that this book, and the accompanying learner and instructor resources, help you to better ponder what clinical questions to ask, to know ways to efficiently find research that answers those questions, to know how to decide whether the results of research are believable, important and applicable, and to use good evidence with patients to provide health care as responsibly, effectively and collaboratively as possible.

Tammy Hoffmann
Sally Bennett
Chris Del Mar

July 2017
LEARNING OBJECTIVES

After reading this chapter, you should be able to:

- Understand more about study designs appropriate for answering questions about effects of interventions
- Generate a structured clinical question about an intervention for a clinical scenario
- Appraise the validity of randomised controlled trials
- Understand how to interpret the results from randomised controlled trials and calculate additional results (such as confidence intervals) where possible
- Describe how evidence about the effects of intervention can be used to inform practice
This chapter focuses on research that can inform us about the effects of interventions. Let us first consider a clinical scenario that will be useful for illustrating the concepts that are the focus of this chapter.

**CLINICAL SCENARIO**

You are working in a community health centre. During a regular team meeting, the general practitioner notes that there are a few people diagnosed with heart failure who are having difficulty with depression. The general practitioner has started them on anti-depressants but believes that psychological support such as cognitive behaviour therapy (CBT) may be effective. She suggests that those involved in providing psychosocial care for these people form a small group to look at the research regarding CBT for people with heart failure who also have depression. In your experience of working with people who have depression and heart failure, you are aware that many of them also have problems with anxiety. You are therefore particularly interested in finding evidence about the effectiveness of CBT programs that have looked at improving participants’ anxiety as well as depression.

This clinical scenario raises several questions about the interventions that might be effective for reducing anxiety and depression among people with heart failure. Is cognitive behaviour therapy alone effective in reducing anxiety and depression in people with heart failure? Does CBT in addition to anti-depressants have a beneficial effect over and above the use of anti-depressants alone? If it is effective, how long does the effect last? Is CBT cost-effective for reducing anxiety and depression for people with heart failure?

As we saw in Chapter 1, clinical decisions are made by integrating information from the best available research evidence with information from our patients, the practice context and our clinical experience. Given that one of the most common information needs in clinical practice relates to questions about the effects of interventions, this chapter will begin by reviewing the role of the study design that is used to test intervention effects before explaining the process of finding and appraising research evidence about the effects of interventions.

**STUDY DESIGNS THAT CAN BE USED FOR ANSWERING QUESTIONS ABOUT THE EFFECTS OF INTERVENTIONS**

There are many different study designs that can provide information about the effects of interventions. Some are more convincing than others in terms of the degree of bias that might be in play given the methods used in the study. From Chapter 2, you will recall that bias is any systematic error in collecting and interpreting data. In Chapter 2, we also introduced the concept of hierarchies of evidence. The higher up the hierarchy that a study design is positioned, in the ideal world, the more likely it is that the study design can minimise the impact of bias on the results of the study. That is why randomised controlled trials (sitting second from the top of the hierarchy of evidence for questions about the effects of interventions) are so commonly recommended as the study design that best controls for bias when testing the effectiveness of interventions. Systematic reviews of randomised controlled trials are located above them (at the top of the hierarchy), because they can combine the results of multiple randomised controlled trials. Combining multiple trials can potentially provide an even clearer picture about the effectiveness of interventions provided they are undertaken rigorously. Systematic reviews are explained in more detail in Chapter 12.

One of the best methods for limiting bias in studies that test the effects of interventions is to have a control group. A control group is a group of participants in the study who should be
as similar in as many ways as possible to the intervention group except that they do not receive
the intervention being studied. Let us first have a look at studies that do not use control groups
and identify some of the problems that can occur.

**Studies that do not use control groups**

*Uncontrolled studies* are studies where the researchers describe what happens when participants
are provided with an intervention, but the intervention is not compared with other interventions.
Examples of uncontrolled study designs are case reports, case series, and before and after studies.
These study designs were explained in Chapter 2. The big problem with uncontrolled studies is
that when participants are given an intervention and simply followed for a period of time with
no comparison against another group, it is impossible to tell how much (if any) of the observed
change is due to the effect of the intervention itself or is due to some other factor or explanation.
There are some alternative explanations for effects seen in uncontrolled studies, and these need
to be kept in mind if you use this type of study to guide your clinical decision making. Some of
the forms of bias that commonly occur in uncontrolled studies are described below.

- **Volunteer bias.** People who volunteer to participate in a study are usually systematically
different from those who do not volunteer. They tend to be more motivated and concerned
about their health. If this is not controlled for, it is possible that the results can make the
intervention appear more favourable (that is, more effective) than it really is. This type of bias
can be controlled for by randomly allocating participants, as we shall see later in this chapter.

- **Maturation.** A participant may change between the time of pre-test (that is, before the
intervention is given) and post-test (after the intervention has finished) as a result of
maturation. For example, consider that you wanted to measure the improvement in fine
motor skills that children in grade 2 of school experienced as a result of a fine-motor-
skill intervention program. If you test them again in grade 3, you will not know if the
improvements that occurred in fine motor skills happened because of natural development
(maturation) or because of the intervention.

- **Natural progression.** Many diseases and health conditions will naturally improve over
time. Improvements that occur in participants may or may not be due to the intervention
that was being studied. The participants may have improved on their own with time, not
because of the intervention.

- **Regression to the mean.** This is a statistical trend that occurs in repeated non-random
experiments, where participants’ results tend to move progressively towards the mean
of the behaviour/outcome that is being measured. This is due not to maturation or
improvement over time, but to the statistical likelihood of someone with high scores not
doing as well when a test is repeated or of someone with low scores being statistically
likely to do better when the test is repeated. Suppose, for example, that you used a
behavioural test to assess 200 children who had attention deficit hyperactivity disorder
and scored their risk of having poor academic outcomes, and that you then provided
the 30 children who had the poorest scores with an intensive behavioural regimen and
medication. Even if the interventions were not effective, you would still expect to observe
some improvement in the children’s scores on the behavioural test when it is next given,
due to regression to the mean. When outliers are repeatedly measured, subsequent values
are less likely to be outliers (that is, they are expected to be closer to the mean value of
the whole group). This always happens, and health professionals who do not expect this to
occur often attribute any improvement that is observed to the intervention. The best way
to deal with the problem of regression to the mean is to randomly allocate participants
to either an experimental group or a control group. The regression to the mean effect can only be accounted for by using a control group (which will have the same regression to the mean if the randomisation succeeded and the two groups are similar). How to determine this is explained later in this chapter.

- **Placebo effect.** This is a well-known type of bias where an improvement in the participants’ condition occurs because they expect or believe that the intervention they are receiving will cause an improvement (even though, in reality, the intervention may not be effective at all).

- **Hawthorne effect.** This is a type of bias that can occur when participants experience improvements not because of the intervention that is being studied, but because of the attention that participants are receiving from being a part of the research process.

- **Rosenthal effect.** This occurs when participants perform better because they are expected to and, in a sense, this expectation has a similar sort of effect as a self-fulfilling prophecy.

**Controlled studies**

It should now be clear that the best way to limit the influence of bias and extraneous factors on the results of a study is to have a control group that can be compared with the intervention group in the study. However, it is not as simple as just having a control group as part of the study. The way in which the control group is created can make an enormous difference to how well the study design actually controls for bias.

**Non-randomised controlled studies**

You will recall from the hierarchy of evidence about the effects of interventions (in Chapter 2) that case-control and cohort studies are located above uncontrolled study designs. This is because they make use of control groups. Cohort studies follow a cohort that has been exposed to a situation or intervention and have a comparison group of people who have not been exposed to the situation of interest (for example, they have not received any intervention). However, because cohort studies are observational studies, the allocation of participants to the intervention and control groups is not under the control of the researcher. It is not possible to tell whether the participants in the intervention and the control groups are similar in terms of all the important factors and, therefore, it is unclear to what extent the exposure (that is, the intervention) might be the reason for the outcome rather than some other factor.

We saw in Chapter 2 that a case-control study is one in which participants with a given disease (or health condition) in a given population (or a representative sample) are identified and are compared with a control group of participants who do not have that disease (or health condition). When a case-control study has been used to answer a question about the effect of an intervention, the ‘cases’ are participants who have been exposed to an intervention and the ‘controls’ are participants who have not. As with cohort studies, because this is an observational study design, the researcher cannot control the assembly of the groups under study (that is, which participants go into which group). Although the controls that are assembled may be similar in many ways to the ‘cases’, it is unlikely that they will be similar with respect to both known and unknown confounders. Chapter 2 explained that confounders are factors that can become confused with the factor of interest (in this case, the intervention that is being studied) and obscure the true results.

In a non-randomised experimental study, the researchers can control the assembly of both experimental and control groups, but the groups are not assembled using random allocation. In non-randomised studies, participants may choose which group they want to be in, or they may
be assigned to a group by the researchers. For example, in a non-randomised experimental study that is evaluating the effectiveness of a particular public health intervention (such as an intervention that encourages walking to work) in a community setting, a researcher may assign one town to the experimental condition and another town to the control condition. The difficulty with this approach is that the people in these towns may be systematically different from each other and so confounding factors, rather than the intervention that is being trialled, may be the reason for any difference that is found between the groups at the end of the study.

So, not only are control groups essential, but in order to make valid comparisons between groups they must also be as similar as possible at the beginning of a study. This is so that we can say, with some certainty, that any differences found between groups at the end of the study are likely to be due to the factor under study (that is, the intervention) rather than because of bias or confounding. To maximise the similarity between groups at the start of a study, researchers need to control for both known and unknown variables that might influence the results. The best way to achieve this is through randomisation. Non-randomised studies are inherently biased in favour of the intervention that is being studied, which can lead researchers to reach the wrong conclusion about the effectiveness of the intervention.²

**Randomised controlled trials**

The key feature of randomised controlled trials is that the participants are randomly allocated to either an intervention (experimental) group or a control group. The outcome of interest is measured in participants in both groups before (known as pre-test) and again after (known as post-test) the intervention has been provided. Therefore, any changes that appear in the intervention group pre-test to post-test, but not in the control group, can reasonably be attributed to the intervention. Figure 4.1 shows the basic design of a randomised controlled trial.

You may notice that we keep referring to how randomised controlled trials can be used to evaluate the effectiveness of an intervention. It is worth noting that they can also be used to evaluate the efficacy of an intervention. **Efficacy** refers to interventions that are tested in ideal circumstances, such as where intervention protocols are very carefully supervised and participant selection is very particular. **Effectiveness** is an evaluation of an intervention in circumstances that are more like real life, such as where there is a broader range of participants included and a typical clinical level of intervention protocol supervision. In this sense, effectiveness trials are more pragmatic in nature (that is, they are accommodating of typical practices) than are efficacy trials.

---

**Figure 4.1**
Basic design of a randomised controlled trial
There are a number of variations on the basic randomised controlled trial design that partly depend on the type or combination of control groups used. There are many variations on what the participants in a control group in a randomised controlled trial actually receive. For example, participants may receive no intervention of any kind (a ‘no intervention’ control), or they may receive a placebo, some form of social control or a comparison intervention. In some randomised controlled trials, there are more than two groups. For example, in one study there might be two intervention groups and one control group; or, in another study, there might be an intervention group, a placebo group and a ‘no intervention’ group. Randomised crossover studies are a type of randomised controlled trial in which all participants take part in both intervention and control groups, but in random order. For example, in a randomised crossover trial of transdermal fentanyl (a pain medication) and sustained-release oral morphine (another pain medication) for treating chronic non-cancer pain, participants were assigned to one of two intervention groups. One group was randomised to four weeks of treatment with sustained-release oral morphine, followed by transdermal fentanyl for four weeks. The second group received the same treatments but in reverse order. A difficulty with crossover trials is that there needs to be a credible wash-out period. That is, the effects of the intervention provided in the first phase must no longer be evident prior to commencing the second phase. In the example we used here, the effect of oral morphine must be cleared prior to the fentanyl being provided.

As we have seen, the advantage of a randomised controlled trial is that any differences found between groups at the end of the study are likely to be due to the intervention rather than to extraneous factors. But the extent to which these differences can be attributed to the intervention is also dependent on some of the specific design features that were used in the trial, and these deserve close attention. The rest of this chapter will look at randomised controlled trials in more depth within the context of the clinical scenario that was presented at the beginning of this chapter. In this scenario, you are a health professional who is working in a small group at a community health centre and you are looking for research regarding the effectiveness of CBT for people with heart failure who also have anxiety and depression. To locate relevant research, you start by focusing on what it is that you specifically want to know about.

**HOW TO STRUCTURE A QUESTION ABOUT THE EFFECT OF AN INTERVENTION**

In Chapter 2, you learnt how to structure clinical questions using the PICO format: Patient/Problem/Population, Intervention/Issue, Comparison (if relevant) and Outcomes.

In our depression and anxiety in heart failure clinical scenario, the population that we are interested in is people who have been diagnosed with heart failure and also have anxiety and depression. We know from our clinical experience and the literature that this sub-group of people with heart failure are at risk of poorer health outcomes.

The intervention that we think could make a difference is cognitive behaviour therapy. This is because you know that CBT has been shown to be effective with people with major depressive disorders, but who do not have heart failure. Are we interested in a comparison intervention? While we could compare the effectiveness of one type of intervention with another, for this scenario it is probably more useful to start by first thinking about whether the intervention is effective over and above usual care. To do this, we would need to compare the intervention to either a placebo (a concept we will discuss later) or to usual care.

There are a number of outcomes that we could consider important for people with heart failure who have depression and anxiety. The most obvious outcome of interest is a reduction in anxiety and depressive symptoms. However, we could also look for interventions that consider objective outcomes such as re-hospitalisation and even mortality. Another outcome (of a more subjective nature) that is likely to be important to people with heart failure is quality of life.
EVIDENCE-BASED PRACTICE ACROSS THE HEALTH PROFESSIONS

CLINICAL SCENARIO (CONTINUED)

THE QUESTION

While there are a number of questions about interventions that can be drawn from the scenario presented at the beginning of this chapter, you decide to form the following clinical question:

In people with heart failure who also have depression and anxiety, is CBT effective in reducing depression and anxiety compared with usual care?

HOW TO FIND EVIDENCE TO ANSWER QUESTIONS ABOUT THE EFFECTS OF AN INTERVENTION

Our clinical scenario question is one about the effectiveness of an intervention for reducing anxiety and depression. You can use the hierarchy of evidence for this type of question as your guide in deciding which type of study you are looking for and where to start searching. In this case, you are looking for a systematic review of randomised controlled trials. If there is no relevant systematic review, you should next look for a randomised controlled trial. If no relevant randomised trials are available, you would then need to look for the next best available type of research, as indicated by the hierarchy of evidence for this question type shown in Chapter 2.

As we saw in Chapter 3, the best source of systematic reviews of randomised controlled trials is the Cochrane Database of Systematic Reviews, so this would be the logical place to start searching. The Cochrane Library also contains the Cochrane Central Register of Controlled Trials, which includes a large collection of citations of randomised trials.

Once you have found a research article that you are interested in, it is important to critically appraise it. That is, you need to examine the research closely to determine whether and how it might inform your clinical practice. As we saw in Chapter 1, when critically appraising research, there are three main aspects to consider: (1) its internal validity (in particular, the risk of bias); (2) its impact (the size and importance of any effect found); and (3) whether or how the evidence might be applicable to your patient or clinical practice.

CLINICAL SCENARIO (CONTINUED)

FINDING EVIDENCE TO ANSWER YOUR QUESTION

You search the Cochrane Database of Systematic Reviews, and find no reviews concerning cognitive behaviour therapy for anxiety and depression in people with heart failure. A search of PubMed found a number of reviews about the treatment of depression in people with heart failure, although they did not include methods important to systematic reviews such as appraisal of the risk of bias in individual studies, or did not focus on cognitive behaviour therapy or on heart failure.

You therefore decide to look for randomised controlled trials in PubMed—Clinical Queries using the search string: depression AND heart failure AND cognitive behaviour therapy (using PubMed—Clinical Queries’ narrow scope function). Your search finds three relevant randomised controlled trials and you choose the one that most closely matches your PICO. It is a randomised controlled trial that has investigated the effectiveness of CBT for depression and self-care in people with heart failure. You obtain the full text of the article in order to critically appraise it, to determine whether the trial also measured anxiety as an outcome, and if it did, to specifically examine the results of the trial and determine whether the findings may be applicable to your clinical scenario.
In this chapter, we will discuss six criteria that are commonly used for appraising the potential risk of bias in a randomised controlled trial. These six criteria are summarised in Box 4.1 (on the following page) and can be found in the *Users’ guides to the medical literature* and in many appraisal checklists such as the Critical Appraisal Skills Program (CASP) checklist and the PEDro scale. A number of studies have demonstrated that estimates of treatment effects may be distorted in trials that do not adequately address these issues. As you work through each of these criteria when appraising an article, it is important to consider the direction of the bias (that is, is it in favour of the intervention or the control group?) as well as its magnitude. As we pointed out in...
Chapter 1, all research has flaws. However, we do not just want to know what the flaws might be, but whether and how they might influence the results of a study.

**Was the assignment of participants to groups randomised?**

Randomised controlled trials, by definition, randomise participants to either the experimental or the control condition. The basic principle of randomisation is that each participant has an equal chance of being assigned to any group, such that any difference between the groups at the beginning of the trial can be assumed to be due to chance. The main benefit of randomisation is related to the idea that, this way, both known and unknown participant characteristics should be evenly distributed between the intervention and control groups. Therefore, any differences between groups that are found at the end of the study are likely to be because of the intervention.\(^{10}\)

Random allocation is best done using a random numbers table, which can be computer-generated. Sometimes it is done by tossing a coin or ‘pulling a number out of a hat’. Additionally, there are different randomisation designs that can be used, and you should be aware of them. Researchers may choose to use some form of restriction, such as blocking or stratification, when allocating participants to groups in order to create a greater balance between the groups at baseline in known characteristics.\(^{11}\) The different randomisation designs are summarised below.

- **Simple randomisation:** involves randomisation of individuals to the experimental or the control condition.
- **Cluster randomisation:** involves random allocation of intact clusters of individuals, rather than individuals (for example, randomisation of schools, towns, clinics or general practices).
- **Stratified randomisation:** in this design, participants are matched and randomly allocated to groups. This method ensures that potentially confounding factors such as age, gender or disease severity are balanced between groups. For example, in a trial that involves people who have had a stroke, participants might be stratified according to their initial stroke severity as belonging to a ‘mild’, ‘moderate’ or ‘severe’ stratum. This way, when randomisation to study groups occurs, researchers can ensure that, within each stratum, there are equal numbers of participants in the intervention and control groups.
- **Block randomisation:** in this design, participants who are similar are grouped into ‘blocks’ and are then assigned to the experimental or control conditions within each block. Block randomisation often uses stratification. An example of block randomisation can be seen in a randomised controlled study of a community-based parenting education intervention.
program designed to increase the use of preventive paediatric healthcare services among low-income, minority mothers.12 Mother–infant dyads \((n = 286)\) who were recruited from four different sites in Washington, DC were assigned to either the standard social services (control) group or the intervention group. To ensure that there were comparable numbers within each group across the four sites, site-specific block randomisation was used. By using block randomisation, selection bias due to demographic differences across the four sites was avoided.

CLINICAL SCENARIO (CONTINUED)

WAS THE ASSIGNMENT OF PARTICIPANTS TO GROUPS RANDOMISED?

In the CBT trial for depression in heart failure, it is explicitly stated that participants were randomised and that the randomisation was carried out using permuted blocks of 2, 4 or 6 pairs stratified by anti-depressant use at baseline.

Was the allocation sequence concealed?

As we have seen, the major benefit of a randomised controlled trial over other study designs is the fact that participants are randomly allocated to the study groups. However, the benefits of randomisation can be undone if the allocation sequence is manipulated or interfered with in any way. As strange as this might seem, a health professional who wants their patient to receive the intervention being evaluated may swap their patient’s group assignment so that their patient receives the intervention being studied. Similarly, if the person who recruits participants to a study knows which condition the participants are to be assigned to, this could influence their decision about whether or not to enrol them in the study. This is why assigning participants to study groups using alternation methods, such as every second person who comes into the clinic, or methods such as date of birth is problematic because the randomisation sequence is known to the people involved.10

Knowledge about which group a participant will be allocated to if they are recruited into a study can lead to the selective assignment of participants, and thus introduce bias into the trial. This knowledge can result in manipulation of either the sequence of groups that participants are to be allocated to or the sequence of participants to be enrolled. Either way, this is a problem. The problem can be dealt with by concealing the allocation sequence from the people who are responsible for enrolling patients into a trial, or from those who assign participants to groups, until the moment of assignment.13 Allocation can be concealed by having the randomisation sequence administered by someone who is ‘off-site’ or at a location away from where people are being enrolled into the study. Another way to conceal allocation is by having the group allocation placed in sealed, opaque envelopes. Opaque envelopes are used so that the group allocation cannot be seen if the envelope is held up to the light! The envelope is not to be opened until the patient has been enrolled into the trial (and is therefore now a participant in the study).

Hopefully, the article that you are appraising will clearly state that allocation was concealed, or that it was done by an independent or off-site person, or that sealed opaque envelopes were used. Unfortunately, though, many studies do not give any indication about whether allocation was concealed,14,15 so you are often left wondering about this, which is frustrating when you are trying to appraise a study. It is possible that some of these studies did use concealed allocation, but you cannot tell this from reading the article.
Were the groups similar at the baseline or start of the trial?

One of the principal aims of randomisation is to ensure that the groups are similar at the start of the trial in all respects, except for whether or not they received the experimental condition (that is, the intervention of interest). However, the use of randomisation does not guarantee that the groups will have similar known baseline characteristics. This is particularly the case if there is a small sample size. Authors of a research article will usually provide data in the article about the baseline characteristics of both groups. This allows readers to make up their own minds as to whether the balance between important prognostic factors (variables that have the potential for influencing outcomes) is sufficient at the start of the trial. Consider, for example, a study about the effectiveness of acupuncture for reducing pain from migraines compared with sham acupuncture. If the participants who were allocated to the acupuncture group had less severe or less chronic pain at the start of the study than the participants who were allocated to the sham acupuncture group, any differences in pain levels that were seen at the end of the study might be the result of that initial difference and not of the acupuncture that was provided.

Differences between the groups that are present at baseline after randomisation have occurred due to chance. Therefore, it is not appropriate to use p values to assess whether these differences are statistically significant. That is, rather than using the p value that is often reported in studies, it is important to examine these differences by comparing means or proportions visually. The extent to which you might be concerned about a baseline difference between the groups depends on how large a difference it is and whether it is a key prognostic variable, both of which require some clinical judgment. The stronger the relationship between the characteristic and the outcome of interest, the more the differences between groups will weaken the strength of any inference about efficacy. For example, consider a study that is investigating the effectiveness of group therapy in improving communication for people who have chronic aphasia following stroke compared with usual care. Typically, such a study would measure and report a wide range of variables at baseline (that is, prior to the intervention), such as participants’ age, gender, education level, place of residence, time since stroke, severity of aphasia, side of stroke, and so on. Some of these variables are more likely to influence communication outcomes than others. The key question to consider is: are any differences in key prognostic variables between the groups large enough that they may have influenced the outcome(s)? Hopefully, if differences are evident the researchers will have corrected for this in the data analysis process.

As a reader (and critical appraiser) of research articles, it is important that you are able to see data for key characteristics that may be of prognostic value in both groups. Many articles will present such data in a table, with the data for the intervention group presented in one column and the data for the control group in another. This enables you to easily compare how similar the groups are for these variables. As well as presenting baseline data about key socio-demographic characteristics (for example, age and gender), articles should also report data about important measures of the severity of the condition (if that is relevant to the study—most times it is) so
that you can see whether the groups were also similar in this respect. For example, in a study that involves participants who have had a stroke, the article may present data about the initial stroke severity of participants, as this variable has the potential to influence how participants respond to an intervention. In most cases, socio-demographic variables alone are not sufficient to determine baseline similarity.

One other area of baseline data that articles should report is the key outcome(s) of the study (that is, the pre-test measurement(s)). To illustrate why this is important, let us consider the example presented earlier of people receiving group communication treatment for aphasia. Although such an article would typically provide information about socio-demographic variables and clinical variables (such as severity of aphasia, type of stroke and side of stroke), having information about participants’ initial (that is, pre-test) scores on the communication outcome measure that the study used would be helpful for considering baseline similarity. This is because, logically, participants’ pre-test scores on a communication measure are likely to be a key prognostic factor for the main outcome of the study, which is communication ability.

When appraising an article, if you do conclude that there are baseline differences between the groups that are likely to be big enough to be of concern, hopefully the researchers will have statistically adjusted for these in the analysis. If they have not, you will need to try to take this into account when interpreting the study.

**CLINICAL SCENARIO (CONTINUED)**

**BASELINE SIMILARITY**

In the CBT trial for depression in heart failure the baseline characteristics shown in Tables 1 and 2 of the article are similar between the study groups (except for a slight difference between groups in racial status) and include most of the likely confounders. Notably, both groups had the same number of participants taking anti-depressants. The baseline scores of the outcome measures were also similar between the two groups. For example, the CBT group had a baseline BDI-II score of 30.7 and the usual care group a BDI-II score of 29.6 (from a possible score range of 0–63).

**Were participants, health professionals and study personnel ‘blind’ to group allocation?**

People involved with a trial, whether they be the participants, the treating health professionals or the study personnel, usually have a belief or expectation about what effect the experimental condition will or will not have. This conscious or unconscious expectation can influence their behaviour, which in turn can affect the results of the study. This is particularly problematic if they know which condition (experimental or control) the participant is receiving. Blinding (also known as ‘masking’) is a technique that is used to prevent participants, health professionals and study personnel from knowing which group the participant was assigned to so that they will not be influenced by that knowledge.

In many studies, it is difficult to achieve blinding. Blinding means more than just keeping the name of the intervention hidden; the experimental and control conditions need to be indistinguishable. This is because even if they are not informed about the nature of the experimental or control conditions (which, for ethical reasons, they usually are) when they sign informed consent forms, participants can often work out which group they are in. Whereas pharmaceutical trials can use placebo medication to prevent participants and health professionals from knowing who
has received the active intervention, blinding of participants and the health professionals who are providing the intervention is very difficult (and often impossible) in many non-pharmaceutical trials. We will now look a little more closely at why it is important to blind participants, health professionals and study personnel to group allocation.

A participant’s knowledge of their treatment status (that is, if they know whether or not they are receiving the intervention that is being evaluated) may consciously or unconsciously influence their performance during the intervention or their reporting of outcomes. For example, if a participant was keen to receive the intervention that was being studied and they were instead randomised to the control group, they may be disappointed and their feelings about this might be reflected in their outcome assessments, particularly if the outcomes being measured are subjective in nature (for example, pain, quality of life or satisfaction). Conversely, if a participant knows or suspects that they are in the intervention group, they may be more positive about their outcomes (such as exaggerating the level of improvement that they have experienced) when they report them, as they wish to be a ‘polite patient’ and are grateful for receiving the intervention.17

The health professionals who provide the intervention often have a view about the effectiveness of interventions, and this can influence the way they interact with the study participants and the way they deliver the intervention. This, in turn, can influence how committed they are to providing the intervention in a reliable and enthusiastic manner, affecting participants’ compliance with the intervention and participants’ responses on outcome measures. For example, if a health professional believes strongly in the value of the intervention that is being studied, they may be very enthusiastic and diligent in their delivery of the intervention, which may in turn influence how participants respond to it. It is easy to see how a health professional’s enthusiasm (or lack of) could influence outcomes. Obviously, some interventions (such as medications) are not able to be influenced easily by the way in which they are provided, but for many other interventions (such as rehabilitation techniques provided by therapists) this can be an issue.

Study personnel who are responsible for measuring outcomes (the assessors) and who are aware of whether the participant is receiving the experimental or control condition may provide different interpretations of marginal findings or differential encouragement during performance tests, either of which can distort results. For example, if an assessor knows that a participant is in the intervention group, they might be a little more generous when scoring a participant’s performance on a task than they would be if they thought that the participant was in the control group. Studies should aim to use blinded assessors to prevent measurement bias from occurring. This can be done by ensuring that the assessor who measures the outcomes at baseline and at follow-up is unaware of the participant’s group assignment. Sometimes this is referred to as the use of an independent assessor. The more objective the outcome that is being assessed, the less critical this issue becomes.10 However, there are not many truly objective outcome measures, as even measures that appear to be reasonably objective (for example, measuring muscle strength manually or functional ability) have a subjective component and, as such, can be susceptible to measurement bias. Therefore, where it is at all possible, studies should try to ensure that the people who are assessing participants’ outcomes are blinded. Estimates from a systematic review reveal that non-blinded assessors (for subjective measures) exaggerated the pooled effect size by 68%.9 Ideally, studies should also check and report on the success of blinding assessors and, where this information is not provided, you may wish to reasonably speculate about whether or not the outcome assessor was actually blinded as claimed.

However, there is a common situation that occurs, particularly in many trials in which non-pharmaceutical interventions are being tested, that makes assessor blinding not possible to achieve. If the participant is aware of their group assignment, then the assessment cannot be considered to be blinded. For example, consider the outcome measure of pain that is assessed using a visual
analogue scale. The participant has to complete the assessment themselves, due to the subjective nature of the symptom experience. In this situation, the participant is really the assessor and, if the participant is not blind to which study group they are in, then the assessor is also not blind to group allocation. Research articles often state that the outcome assessors were blind to group allocation. Most articles measure more than one outcome, and often a combination of objective and subjective outcome measures is used. So, while this statement may be true for objective outcomes, if the article involved outcomes that were assessed by participant self-report and the participants were not blinded, you cannot consider that these subjective outcomes were measured by a blinded assessor.

**CLINICAL SCENARIO (CONTINUED)**

**BLINDING**

In the CBT trial for depression in heart failure, it was not possible to blind participants to which group (CBT or usual care) they had been allocated to, nor to blind therapists to the group allocation. The primary outcome measure was the BDI-II depression score, and this was measured by participant self-report. As the primary outcome measure was measured by participant self-report and participants were not blinded, technically the measurement of the primary outcome measure was not done by a blinded assessor. However, participant self-report is an appropriate, and widely used and accepted, measure of collecting data about depression and anxiety and there is not really any other feasible method that enables this type of data to be collected in a blinded manner. However, the authors state that the outcome assessors were blinded to group assignments. Blinding is certainly possible to achieve for the objective measurements taken in this trial, such as the 6-minute walk test.

**Were all participants who entered the trial properly accounted for at its conclusion, and how complete was follow-up?**

In randomised trials, it is common to have missing data at follow-up. There are many reasons why data may be missing. For example, some questionnaires may not have been fully completed by participants, some participants may have decided to leave the study, or some participants may have moved house and cannot be located at the time of the follow-up assessment. How much of a problem this is for the study, with respect to the bias that is consequently introduced, depends on how many participants left the study and why.

It is therefore helpful to know whether all the participants who entered the trial were properly accounted for. In other words, we want to know what happened to them. Could the reason that they dropped out of the study have affected the results? This may be the case, for example, if they left the study because the intervention was making them worse or causing adverse side effects. If this was the case, this might make the intervention look more effective than it really was. Did they leave the study simply because they changed jobs and moved away, or was the reason that they dropped out related to the study or to their health? For example, it may not be possible to obtain data from participants at follow-up measurement points because they became unwell, or maybe because they improved and no longer wanted to participate. Hopefully, you can now see why it is important to know why there are missing data for some participants.

The more participants who are ‘lost to follow-up’, the more the trial may be at risk of bias because participants that leave the study are likely to have different prognoses from those who stay. It has been suggested that ‘[r]eaders can decide for themselves when the loss to follow-up is
excessive by assuming, in positive trials [that is, trials that showed that the intervention was effective], that all patients lost from the treatment group did badly, and all lost from the control group did well, and then recalculating the outcomes under these assumptions. If the conclusions of the trial do not change, then the loss to follow-up was not excessive. If the conclusions would change, the strength of inference is weakened (that is, less confidence can be placed in the study results).

When large numbers of participants leave a study, the potential for bias is enhanced. Various authors suggest that if more than 15–20% of participants leave the study (with no data available for them), then the results should be considered with much greater caution. Therefore, you are looking for a study to have a minimum follow-up rate of at least 80–85%. To calculate the loss to follow-up, you divide the number of participants included in the analysis at the time point of interest (such as the 6-month follow-up) by the number of participants who were originally randomised into the study groups. This gives the percentage of participants who were followed up. Some articles will clearly report this, while in others it can be straightforward to find the necessary data that you need to calculate this, particularly if the article has provided a flow diagram (see Figure 4.2). It is highly recommended that trials do this, and this is explained more fully later in the chapter when the recommended reporting of a randomised trial is discussed. In other articles, this information can be obtained from the text, typically in the results section. In some articles, the only place to locate information about the number of participants who remain in the groups at a particular time point is from the column headers in a results table. And finally, in some articles, despite all of your best hunting efforts, there may be no information about the number of participants who were retained in the study. This may mean that there were no participants lost to follow-up (which is highly unlikely, as at least some participants are lost to follow-up in most studies) or the authors of the article did not report the loss of participants that occurred. Either way, as you cannot determine how complete the follow-up was, you should be suspicious of the study and consider the results to be potentially biased.

It is worth noting that where there is loss to follow-up in both the intervention and the control groups and the reasons for these losses are both known and similar between these groups, it is less likely that bias will be problematic. When the reasons why participants leave the study are unknown, or when they are known to be different between the groups and potentially prognostically relevant, you should be more suspicious about the validity of the study. That is why it is important to consider the reasons for loss to follow-up and whether the number of participants lost to follow-up was approximately the same for both of the study groups, as well as the actual percentage of participants who were lost to follow-up.

CLINICAL SCENARIO (CONTINUED)

FOLLOW-UP OF PARTICIPANTS

In the CBT trial for depression in heart failure, participants were followed from baseline to 12 months post-randomisation. The flow of participants through the trial is clearly provided in Figure 1 of the article, which also lists the reasons for some people being unable to be followed up. Before the 6-month assessment, 16% of participants were lost to follow-up—that is, an 84% follow-up rate—which is around the 85% follow-up rate that is considered acceptable for trials. There were equal numbers lost from the two groups.

Loss to follow-up at the 9- and 12-month time points measures was higher. This trial used multiple imputation models to impute missing data so that analysis could use a full data set, consistent with analysis by intention-to-treat, which will now be examined.
Were participants analysed in the groups to which they were randomised using intention-to-treat analysis?

The final criterion for assessing a trial’s risk of bias is whether data from all participants were analysed in the groups to which participants were initially randomised, regardless of whether they ended up receiving the treatment. This analysis principle is referred to as intention-to-treat analysis. In other words, participants should be analysed in the group that corresponds to how they were intended to be treated, not how they were actually treated.

It is important that an intention-to-treat analysis is performed, because study participants may not always receive the intervention or control condition as it was allocated (that is, intended to be received). In general, participants may not receive the intervention (even though they were allocated to the intervention group) because they are either unwell or unmotivated, or for other reasons.
reasons related to prognosis.5 For example, in a study that is evaluating the effect of a medication, some participants in the intervention group may forget to take the medication and therefore do not actually receive the intervention as intended. In a study that is evaluating the effects of a home-based exercise program, some of the participants in the intervention group may not practise any of the exercises that are part of the intervention because they are not very motivated. Likewise, in a study that is evaluating a series of small-group education sessions for people who have had a heart attack, some participants in the intervention group may decide not to attend some or all education sessions because they feel unwell. From these examples, you can see that even though these participants were in the intervention group of these studies, they did not actually receive the intervention (either at all or only partly).

It may be tempting for the researchers who are conducting these studies to analyse the data from these participants as if they were in the control group instead. However, doing this would increase the numbers in the control group who were either unmotivated or unwell. This would make the intervention appear more effective than it actually is, because there would be a greater number of participants in the control group who were likely to have unfavourable outcomes. It may also be tempting for researchers to discard the results from participants who did not receive the intervention (or control condition) as was intended. This is also an unsuitable way of dealing with this issue, as these participants would then be considered as lost to follow-up and we saw in the previous criterion why it is important that as few participants as possible are lost to follow-up.

For the sake of completeness, it is important to point out that it is not only participants who are allocated to the intervention group but do not receive the intervention that we should think about. The opposite can also happen. Participants who are allocated to the control group can inadvertently end up receiving the intervention. Again, intention-to-treat analysis should be used, and these participants should still be analysed as part of the control group.

The value of intention-to-treat analysis is that it preserves the value of randomisation. It helps to ensure that prognostic factors that we know about, and those that we do not know about, will still be, on average, equally distributed between the groups. Because of this, any effect that we see, such as improvement in participants’ outcomes, is most likely to be because of the intervention rather than unrelated factors.

The difficulty in carrying out a true intention-to-treat analysis is that the data for all participants are needed. However, as we saw in the previous criterion about follow-up of participants, this is unrealistic to expect and most studies have missing data. There is currently no real agreement about the best way to deal with such missing data, but researchers may sometimes estimate or impute data.20,21 Data imputation is a statistical procedure that substitutes missing data in a data file with estimated data. Other studies may simply report that they have carried out an intention-to-treat analysis or that participants received the experimental or control conditions as allocated without providing details of what was actually done or how missing data were dealt with. In this case, as the reader of the article, you may choose to accept this at face value or to remain sceptical about how this issue was dealt with, depending on what other clues are available in the study report.

**CLINICAL SCENARIO (CONTINUED)**

**INTENTION-TO-TREAT ANALYSIS**

In the CBT trial for depression in heart failure, it is stated that data were analysed using intention-to-treat; and details about how missing data were imputed are provided in the statistical analysis section.
The role of chance

So far in this chapter, we have considered the potential for bias in randomised trials. Another aspect that is important to consider is the possibility that chance might be an alternative explanation for the findings. So, a further question that you may wish to consider when appraising a randomised trial is: did the study report a power calculation that might indicate what sample size would be necessary for the study to detect an effect if the effect actually exists? As we saw in Chapter 2, having an adequate sample size is important so that the study can avoid a type II error occurring. You may remember that a type II error is the failure to find and report a relationship when a relationship actually exists.22

CLINICAL SCENARIO (CONTINUED)

DID THE STUDY HAVE ENOUGH PARTICIPANTS TO MINIMISE THE PLAY OF CHANCE?

In the CBT trial for depression in heart failure, 158 participants were recruited. A power analysis had been performed. It had estimated that 240 participants would be needed to detect a between-group difference of 3 or more points on the BDI-II, previously determined as a clinically significant difference. The authors acknowledge the inability to recruit sufficient numbers but noted that this did not obscure the effect of the intervention on depression. However, small sample sizes (low power) can exaggerate the estimate of the magnitude of the effect23 and this needs to be considered when interpreting the results.

COMPLETENESS OF REPORTING OF RANDOMISED CONTROLLED TRIALS

As we have seen in many places throughout this chapter, it can often be difficult for readers of research studies to know whether a study has met enough requirements to be considered a well-designed and well-conducted randomised trial that is relatively free of bias. To help overcome this problem and aid in the critical appraisal and interpretation of trials, an evidence-based initiative known as the CONSORT (Consolidated Standards of Reporting Trials) statement13,24 has been developed to guide authors in how to completely report the details of a randomised controlled trial. The CONSORT statement is considered an evolving document and, at the time of writing, the most recent checklist contained 25 items (see Figure 4.2). Full details are available at www.consort-statement.org. The CONSORT statement is also used by peer reviewers of articles when the articles are being considered for publication, and many journals now insist that reports of randomised trials follow the CONSORT statement. The CONSORT statement also has a number of approved extensions—such as for particular trial designs (for example, cluster trials),25 interventions (for example, Template for Intervention Description and Replication (TIDieR),26 and types of data (such as harms data).27 This is helping to improve the quality of reporting of trials but, unfortunately, many articles still do not contain all of the information that readers require. Some of the reasons are that not all journals require, or enforce, adherence to the CONSORT statement and older articles were published prior to CONSORT’s development.

After appraising and determining that an article about the effects of an intervention appears to be reasonably free of bias, you then proceed to looking at the importance of the results.

UNDERSTANDING RESULTS

One of the fundamental concepts that you need to keep in mind when trying to make sense of the results of a randomised controlled trial is that clinical trials provide us with an estimate of
the *average* effects of an intervention. Not every participant in the intervention group of a randomised trial is going to benefit from the intervention that is being studied—some may benefit a lot, some may benefit a little, some may experience no change, and some may even be worse (a little or a lot) as a result of receiving the intervention. The results from all participants are combined and the *average* effect of the intervention is what is reported.

Before getting into the details about how to interpret the results of a trial, the first thing that you need to look at is whether you are dealing with continuous or dichotomous data:

- **Variables with continuous data** can take any value along a continuum within a defined range. Examples of continuous variables are age, range of motion in a joint, walking speed, and score on a visual analogue scale.

- **Variables with dichotomous data** have only two possible values. For example, male/female, satisfied/not satisfied with treatment, and hip fracture/no hip fracture.

The way that you make sense of the results of a randomised trial depends on whether you are dealing with outcomes that were continuous or dichotomous. We will look at continuous data first. However, regardless of whether the results of the study were measured using continuous or dichotomous outcomes, we will be looking at how to answer two main questions:

1. What is the **size** of the intervention effect?
2. What is the **precision** of the intervention effect?

**Continuous outcomes—size of the intervention effect**

When you are trying to work out how much of a difference the intervention made, you are attempting to determine the size of the intervention effect. When you are dealing with continuous data, this is often quite a straightforward process.

The best estimate for the size of the intervention effect is the **difference in means** (or medians, if that is what is reported) *between the intervention and the control groups*. Many articles will already have done this for you and will report the difference. In other articles, you will have to do the simple subtraction calculation yourself. Figure 4.3 shows a hypothetical example in which the mean pain (on a 0–100 scale) experienced by the control group is higher (80 points) than the mean pain for the intervention group (50 points) when measured post-intervention. The effect size is simply the difference between these two means (80 − 50 = 30 points). Those in the intervention group had (on average) 30 points less than those in the control group.

Let us consider a real example from a randomised controlled trial. In a trial of the efficacy of a self-management program for people with knee osteoarthritis in addition to usual care, compared with usual care, one of the main outcome measures was pain, which was measured using a 0–10 cm visual analogue scale with 0 representing ‘no pain at all’ and 10 representing ‘the worst pain imaginable’. At the 3-month follow-up, the mean reduction in knee pain was 0.67 cm (standard deviation [SD] = 2.10) in the intervention group and 0.01 cm (SD = 2.00) in the control group. This difference was statistically significant (*p* = 0.023). You can calculate the intervention effect size (difference in mean change between the groups) as: 0.67 cm minus 0.01 cm = 0.66 cm.

Note that, in this study, the authors reported the mean *improvement* (*reduction in pain*) scores at the 3-month follow-up (that is, the *change* in pain from baseline to the 3-month follow-up). Some studies report *change scores*; others report *end scores* (which are the scores at the end of the intervention period or follow-up period). Regardless of whether change scores or end scores are reported, the method of calculating the size of the intervention effect is the same. However, when dealing with change scores, you need the difference of the mean *change* between the intervention and control groups.
Clinical significance

Once you know the size of the intervention effect, you need to decide whether this result is clinically significant. As we saw in Chapter 2, just because a study finds a statistically significant result, it does not mean that the result is clinically significant. Deciding whether a result is clinically significant requires your judgment (and, ideally, your patient’s, too) on whether the benefits of the intervention outweigh its costs. ‘Costs’ should be regarded in the broadest sense to be any inconveniences, discomforts or harms associated with the intervention, in addition to any monetary costs. To make decisions about clinical significance, it helps to determine what the smallest intervention effect is that you and your patient consider to be clinically worthwhile. Where possible, this decision is one that should be reached in conjunction with the patient so that their preferences are considered. (This topic is considered further in Chapter 14.)

To interpret the results of a study, the smallest clinically worthwhile difference (to warrant using a self-management program in addition to standard care) needs to be established. Many trials do not comment on the clinical significance of the results, and when they do, it is often mentioned without any elaboration or justification.\(^2^9\) This might be decided directly by the health professional, by using guidelines established by research on the particular measure being used (if available), by consultation with the patient, or by some combination of these approaches. Frequently, health professionals make a clinical judgment based on their experience with a particular measure and by subsequent discussion with the patient about their preferences in relation to the costs (including both financial costs and inconveniences) involved. Health professionals then need to consider how the smallest clinically worthwhile difference that was determined relates to the effect found in the study. This determination is handled in a number of different ways in the literature.
One of the earliest methods for deciding important differences using effect sizes was developed by Cohen.\(^{30}\) Effect sizes (represented by the symbol \(d\)) were calculated by taking the difference between group average scores and dividing it by the average of the standard deviation for both groups. This effect size is then compared with ranges classified intuitively by Cohen: 0.2 being a small effect size, 0.5 a moderate effect size and 0.8 a large effect size. This general rule of thumb has consequently been used to determine whether or not a change or difference was important.

Norman and colleagues looked at effect sizes empirically (using a systematic review of 62 effect sizes from 38 studies of chronic disease that had calculated minimally important differences) and found that, in most circumstances, the smallest detectable difference was approximately half a standard deviation.\(^{31}\) They thus recommend that, when no other information exists about the smallest clinically worthwhile difference, an intervention effect should be regarded as worthwhile if it exceeds half a standard deviation. However, the use of statistical standards for interpreting differences has been criticised because the between-patient variability, or number of standard deviation units, depends on the heterogeneity of the population (meaning whether the population is made up of different sorts of participants).\(^ {32}\)

A more direct approach is to simply compare the mean intervention effect with a nominated smallest clinically worthwhile difference. If the mean intervention effect lies below the smallest clinically worthwhile difference, we may consider it to be not clinically significant. For our knee osteoarthritis example, let us assume that in conjunction with our patient we nominate a 20% reduction of initial pain as the smallest difference that we would consider to be clinically worthwhile in order to add self-management to what the patient is already receiving. Our calculations show that a 20% reduction from the initial average pain of 4.05 cm experienced by participants in this study would be 0.8 cm. If the intervention has a greater effect than a 20% reduction in pain from baseline, we may consider that the benefits of it outweigh the costs and may therefore use it with our patient(s). In this study, the difference between groups in mean pain reduction (0.66 cm) is less than 0.8 cm, so we might be tempted to conclude that the result may not be clinically significant for this patient.

An alternative approach that is sometimes used for considering the effect size relative to baseline values involves comparing the effect size to the scale range of the outcome measure. In this example, we would simply compare the intervention effect size of 0.66 cm in relation to the overall possible score range of 0–10 cm, and see that this between-group difference is not very large and therefore conclude that the result is unlikely to be considered clinically significant.

Note, however, that as this is an average effect there may be some patients who do much better than this and, for them, the intervention is clinically significant. Of course, conversely, there may be some patients who do much worse. This depends on the distribution of changes that occur in the two groups. One way of dealing with this is to look for the proportion of patients in both groups who improved, stayed the same or got worse relative to the nominated smallest clinically worthwhile difference. However, these data are often elusive, as they are often not reported in articles.

Another approach to determining clinical significance takes into account the uncertainty in measurement using the confidence intervals around the estimate. To understand this approach, we need to first look at confidence intervals in some detail.

**Continuous outcomes—precision of the intervention effect**

**How are confidence intervals useful?**

At the beginning of this results section, we highlighted that the results of a study are only an estimate of the true effect of the intervention. The size of the intervention effect in a study
approximates but does not equal the true size of the intervention effect in the population represented by the study sample. As each study only involves a small sample of participants (regardless of the actual sample size, it is still just a sample of all of the patients who meet the study's eligibility criteria and therefore is small in the grand scheme of things), the results of any study are just an estimate based on the sample of participants in that particular study. If we replicated the study with another sample of participants, we would (most likely) obtain a different estimate. As we saw in Chapter 2, the true value refers to the population value, not just the estimate of a value that has come from the sample of one study.

Confidence intervals are a way of describing how much uncertainty is associated with the estimate of the intervention effect (in other words, the precision or accuracy of the estimate). We saw in Chapter 2 how confidence intervals provide us with a range of values within which the true value lies. When dealing with 95% confidence intervals, what you are saying is that you are 95% certain that the true average intervention effect lies between the upper and the lower limits of the confidence interval. In the knee osteoarthritis trial that we considered above, the 95% confidence interval for the difference of the mean change is 0.05 cm to 1.27 cm. (See Box 4.2, on the following page, for how this was calculated.) So, we are 95% certain that the true average intervention effect (at 3 months follow-up) of the self-management program on knee pain in people with osteoarthritis lies between 0.05 cm and 1.27 cm.

**How do I calculate a confidence interval?**

Hopefully, the study that you are appraising will have included confidence intervals with the results in the results section. Fortunately, this is becoming increasingly common in research articles (probably because of the CONSORT statement and growing awareness of the usefulness of confidence intervals). If not, you may be able to calculate the confidence interval if the study provides you with the right information to do so. (See Box 4.2 for what you need.) An easy way of calculating confidence intervals is to use an online calculator. There are plenty available, such as those found at www.pedro.org.au/english/downloads/confidence-interval-calculator/. If the internet is not handy, you can use a simple formula to calculate the confidence interval for the difference between the means of two groups. (The formula is shown in Box 4.2.)

**Confidence intervals and statistical significance**

Confidence intervals can also be used to determine whether a result is statistically significant. Consider a randomised trial where two interventions were being evaluated and, at the end of the trial, it was found that, on average, participants in both groups improved by the same amount. If we were to calculate the size of the intervention effect for this trial it would be zero, as there would be no (0) difference between the means of the two groups. When referring to the difference between two groups with means, zero is considered a ‘no effect’ value. Therefore,

- if a confidence interval includes the ‘no effect’ value, the result is not statistically significant; and the opposite is also true:
- if the confidence interval does not include the ‘no effect’ value, the result is statistically significant.

In the knee osteoarthritis trial, we calculated the 95% confidence interval to be 0.05 cm to 1.27 cm. This interval does not include the ‘no effect’ value of zero, so we can therefore conclude that the result is statistically significant without needing to know the p value; although, in this case, the p value (0.023) was provided in the article and it also indicates a statistically significant result.

As an aside, if a result is not statistically significant, it is incorrect to refer to this as a ‘negative’ difference and to imply that the study has shown no difference and conclude that the intervention
was not effective. It has not done this at all. All that the study has shown is an absence of evidence of a difference. A simple way to remember this is that non-significance does not mean no effect.

Confidence intervals and clinical significance

We now return to our previous discussion about clinical significance. Earlier we saw that there are a number of approaches that can be used to compare the effect estimate of a study to the smallest clinically worthwhile difference that is established by the health professional (and sometimes their patient as well). We will now explain a useful way of considering clinical significance that involves using confidence intervals to help make this decision.

Before we go on to explain the relationship between confidence intervals and clinical significance, you may find it easier to understand confidence intervals by viewing them on a tree plot (see Figure 4.4). A tree plot is a line along which varying intervention effects lie. The ‘no effect’ value

---

**How to calculate the confidence interval (CI) for the difference between the means of two groups**

A formula that can be used is:

\[
95\% \text{ CI} = \text{Difference} \pm 3 \times \frac{\text{SD}}{\sqrt{n_{av}}}
\]

where:
- Difference = difference between the two means
- SD = average of the two standard deviations
- \( n_{av} \) = average of the group sizes

For the knee osteoarthritis study, Difference = 0.66; SD = \((2.10 + 2.00) ÷ 2 = 2.05\); and \( n_{av} = (95 + 107) ÷ 2 = 101\).

Therefore:

\[
95\% \text{ CI} = 0.66 \pm 0.61
\]

\[
= 0.05 \text{ to } 1.27
\]

When you calculate confidence intervals yourself, they will vary slightly depending on whether you use this formula or an online calculator. This formula is an approximation of the complex equation that researchers use to calculate confidence intervals for their study results, but it is adequate for the purposes of health professionals who are considering using an intervention in clinical practice and wish to obtain information about the precision of the estimate of the intervention’s effect.

Occasionally, you might calculate a confidence interval that is at odds with the \( p \) value reported in the paper (that is, the confidence interval might indicate non-significance when in fact the \( p \) value in the paper is significant). This might occur because the test used by the researchers does not assume a normal distribution (as the 95% confidence interval does) or because the \( p \) value was close to 0.05 and the rough calculation of the confidence interval might end up including zero as it is a less precise calculation.

Note: if the study reports standard errors (SEs) instead of standard deviations, the formula to calculate the confidence interval is:

\[
95\% \text{ CI} = \text{Difference} \pm 3 \times \text{SE}
\]

is indicated in Figure 4.4 as the value 0. Effect estimates to the left of the ‘no effect’ value may indicate harm. Also marked on the figure is a dotted line that indicates the supposed smallest clinically worthwhile intervention effect. Anything to the left of this line but to the right of the ‘no effect’ value estimate represents effects of the intervention that are too small to be worthwhile. Conversely, anything to the right of this line indicates intervention effects that are clinically worthwhile.

In the situation where the entire confidence interval is below the smallest clinically worthwhile effect, this is a useful result. It is useful because at least we know with some certainty that the intervention is not likely to produce a clinically worthwhile effect. Similarly, when an entire confidence interval is above the smallest clinically worthwhile effect, this is a clear result, as we know with some certainty that the intervention is likely to produce a clinically worthwhile effect.

However, in the knee osteoarthritis trial, we calculated the 95% confidence interval to be 0.05 cm to 1.27 cm. The lower value is below 0.8 cm (20% of the initial pain level that we nominated as the smallest clinically worthwhile effect), but the upper value is above 0.8 cm. We can see this clearly if we mark the confidence interval on to a tree plot (see Figure 4.5, tree plot A, on the following page). CI range indicates that there is uncertainty about whether or not a clinically worthwhile effect is occurring.

When the confidence interval spans the smallest clinically worthwhile effect, it is more difficult to interpret clinical significance. In this situation, the true effect of the intervention could lie either above the smallest clinically worthwhile effect or below it. In other words, there is a chance that the intervention may produce a clinically worthwhile effect, but there is also a chance that it may not. Another example of this is illustrated in tree plot B in Figure 4.5, using the data from a randomised controlled trial that investigated the efficacy of a guided self-management program for people with asthma compared with traditional asthma treatment.35 One of the main outcome measures in this study was quality of life, which was measured using a section of the St George Respiratory Questionnaire. The total score for this outcome measure can range from −50 to +50, with positive scores indicating improvement and negative scores indicating deterioration in quality of life compared with one year ago. At the 12-month follow-up, the difference between the means of the two groups (that is, the intervention effect size) was 8 points (in favour of the
EVIDENCE-BASED PRACTICE ACROSS THE HEALTH PROFESSIONS

self-management group), with a 95% confidence interval of 2 to 15. This difference was statistically significant \((p = 0.009)\). Let us assume that we nominate a difference of 4 points to be the smallest clinically worthwhile effect for this example. In this example, the mean difference (8) is above what we have chosen as the smallest clinically worthwhile effect (4), but the confidence interval includes some values that are above the worthwhile effect and some values that are below it. If the true intervention effect was at the upper limit of the confidence interval (at 15), we would consider the intervention to be worthwhile, while if the true intervention effect was at the lower limit of the confidence interval (2), we may not. So, although we would probably conclude that

---

**Figure 4.5**

Tree plots showing the effect size, the smallest clinically worthwhile effect and the confidence interval associated with the effect size

In tree plot A, the estimate of the intervention effect size (0.66 cm) sits below the smallest clinically worthwhile effect of 0.8 cm and the confidence interval (0.05 cm to 1.27 cm) spans the smallest clinically worthwhile effect of 0.8 cm. The estimate of the intervention effect size is indicated as a small square, the 95% confidence interval about this estimate is shown as a horizontal line, and the dotted line indicates the supposed smallest clinically worthwhile intervention effect. In tree plot B, the estimate of the intervention effect size (8) is above the smallest clinically worthwhile effect (4) and the confidence interval (2 to 15) spans the smallest clinically worthwhile effect.
the effect of this self-management intervention on quality of life was clinically significant, this conclusion would be made with a degree of uncertainty.

The situation of a confidence interval spanning the smallest clinically worthwhile effect is a common one, and there are two main reasons why it can occur. First, it can occur when a study has a small sample size and therefore low power. The concept of power was explained in Chapter 2. As we also saw in Chapter 2, the smaller the sample size of a study, the wider (that is, less precise) the confidence interval is, which makes it more likely that the confidence interval will span the worthwhile effect. Second, it can occur because many interventions only have fairly small intervention effects, meaning that their true effects are close to the smallest clinically worthwhile effect. As a consequence, they need to have very narrow confidence intervals if the confidence interval is going to avoid spanning the smallest worthwhile effect. As we just discussed, this typically means that a very large sample size is needed. In allied health studies in particular, this can be difficult to achieve.

There are two ways that you, as a health professional who is trying to decide whether to use an intervention with a patient, can deal with this uncertainty:

- Accept the uncertainty and make your decision according to whether the difference between the group means is higher or lower than the smallest clinically worthwhile effect. However, keep the confidence interval in mind, as it indicates the degree of doubt that you should have about this estimate.

**Clinical Scenario (continued)**

**Main Results—Depression**

In the CBT trial for depression in heart failure, depressive symptoms measured by the BDI-II are reported as means (and standard deviations) at each time point. At 6 months (the main time point of interest in this trial), the mean difference between the two groups was $-4.43$ points ($95\% \text{ CI } -7.68 \text{ to } -1.18$), which was statistically significant. There are a few points to note about this. First, the negative sign indicates a reduction in depression (because, on this outcome measure, lower scores indicate less severe depression). Second, the confidence interval does not contain zero, and therefore this result was statistically significant.

Understanding the clinical significance of this result is more complex. At the time of writing, no research was found that specifically considers the smallest clinically worthwhile effect for the BDI-II for people with heart disease. The National Institute for Health and Care Excellence (NICE), however, suggests that a difference of 3 or more BDI-II points is a clinically significant treatment effect for depression in the general population. On this basis, the depressive symptoms in participants in both groups improved, on average, by a clinically significant amount; however, it is the between-group difference (a mean of $-4.43$) that we are really interested in, and it is also larger than the 3 points identified by the NICE guidelines. This suggests that this effect might be considered clinically significant, although this might ultimately be determined by discussion between the clinician and each patient. It also should be noted that the confidence interval for this mean change spans the 3 point score ($-7.68 \text{ to } -1.18$), meaning that the effect may be as small as a 1.18 reduction for some patients, which some may consider to be not clinically significant.

On the Beck Anxiety Inventory, the mean difference between the two groups at 6 months was $-4.47$ ($95\% \text{ CI } -7.70 \text{ to } -1.25$), which is statistically significant as the confidence interval does not contain zero. As the literature does not suggest what the smallest clinically worthwhile effect might be (particularly for people with heart disease), we could talk with our patients to see what amount of reduction in anxiety they would want to achieve.
• Try to increase the certainty by searching for similar studies and establishing whether the findings are replicated in other studies. This is one of the advantages of a systematic review, in particular a meta-analysis, as combining the results from multiple trials increases the sample size. The consequence of this is usually a more narrow (more precise) confidence interval which is less likely to span the smallest clinically worthwhile effect. Systematic reviews and meta-analyses are discussed in detail in Chapter 12.

Dichotomous outcomes—size of the treatment effect

Often the outcomes that are reported in a trial will be presented as dichotomous data. As we saw at the beginning of this results section, these are data for which there are only two possible values. It is worth being aware that data measured using a continuous scale can also be categorised, using a certain cut-off point on the scale, so that the data become dichotomised. For example, data on a 10-point visual analogue pain scale can be arbitrarily dichotomised around a cut-off point of 3 (or any point that the researchers choose), so that a pain score of 3 or less is categorised as mild pain and a score of above 3 is categorised as moderate/severe pain. By doing this, the researchers have converted continuous data into data that can be analysed as dichotomous data.

Health professionals and patients are often interested in comparative results—the outcome in one group relative to the outcome in the other group. This overall (comparative) consideration is one of risk. Before getting into the details, let us briefly review the concept of risk. Risk is simply the chance, or probability, of an event occurring. A probability can be described by numbers, ranging from 0 to 1, and is a proportion or ratio. Risks and probabilities are usually expressed as a decimal, such as 0.1667, which is the same as 16.67%. Risk can be expressed in various ways (but using the same data).37 We will consider each of these in turn:

1. relative risk (or the flip side of this, which is referred to as ‘relative benefit’)
2. relative risk reduction (or relative benefit increase)
3. absolute risk reduction (or absolute benefit increase)
4. number needed to treat.

Risk and relative risk (or relative benefit)

Consider a hypothetical study that investigated an exercise intervention to prevent the occurrence of falls in older adults. The control group \((n = 100)\) received no intervention and is compared with the intervention group \((n = 100)\) who did not receive the exercise intervention. Suppose that at the end of the 6-month trial, 25 of the participants in the control group had had a fall.

- The risk for occurrence of a fall in the control group can be calculated as \(25/100\), which can be expressed as 25%, or a risk of occurrence of 0.25. This is also known as the ‘Control Event Rate’ (CER).
- If, in the exercise group, only 5 participants had had a fall, the risk of occurrence would be \(5/100\), or 0.05. This is sometimes referred to as the ‘Experimental Event Rate’ (EER).

These data can be represented graphically, as in Figure 4.6.

We are interested in the comparison between groups. One way of doing this is to consider the relative risk—a ratio of the probability of the event occurring in the intervention group versus in the control group.38 In other words, relative risk is the risk or probability of the event in the intervention group divided by that in the control group. The term ‘relative risk’ is being increasingly replaced by ‘risk ratio’, which describes the same concept but reflects that it is a ratio that is being referred to. In the exercise for falls study, the relative risk (that is, the risk ratio) would be calculated by dividing the EER by the CER:
This can be expressed as: ‘The risk of having a fall for those in the exercise group is 20% of the risk in the control group.’

You may remember that when we were discussing continuous outcomes, the ‘no effect’ value was zero (when referring to the difference between group means). When referring to relative risk, the ‘no effect’ value (the value that indicates that there is no difference between groups) is 1. This is because if the risk is the same in both groups (for example, 0.25 in one group and 0.25 in the other group), there would be no difference and the relative risk (risk ratio) would be $0.25/0.25 = 1$. Therefore, a relative risk (risk ratio) of less than 1 indicates lower risk—that is, a benefit from the intervention.38

If we are evaluating a study that aimed to improve an outcome (that is, make it more likely to occur) rather than reducing a risk, we might instead consider using the equivalent concept of relative benefit. For example, in a study that evaluated the effectiveness of phonological training to increase reading accuracy among children with learning disabilities, we could calculate relative benefit as the study was about improving an outcome, not reducing the risk of it happening.

Sometimes a study will report an odds ratio instead of relative risk. It is similar, except that an odds ratio refers to a ratio of odds, rather than a ratio of risks. The odds ratio is the ratio of the odds of an event for those in the intervention group compared with the odds of an event in the control group. Odds are derived by dividing the event rate by the non-event rate for each group. The odds ratio is calculated by the following formula, where, again, ‘CER’ refers to the Control Event Rate and ‘EER’ to the Experimental Event Rate:

$$\text{Odds ratio (OR)} = \frac{\text{EER}/(1 - \text{EER})}{\text{CER}/(1 - \text{CER})}$$

Also, some studies may report the intervention effect as a hazard ratio. This typically occurs when a survival analysis has been performed. A hazard ratio is broadly similar in concept to a risk ratio (relative risk) and so can be interpreted in a similar manner—that is, it describes how many times more (or less) likely a participant is to experience the event at a particular point in time if they were in the intervention group rather than the control group.

**Relative risk reduction (or relative benefit increase)**

Relative risk reduction is another way of expressing the difference between the two groups, and is simply the proportional reduction in an event of interest (for example, falls) in the intervention group compared with the control group, at a specified time point. Again, there is a flip side of this concept, which is known as ‘relative benefit increase’. When appraising an article and using this concept, you need to keep in mind whether you are considering negative or positive outcomes:
• Relative risk reduction (RRR) is used to express the reduction in risk of a negative outcome (such as falling).
• Relative benefit increase (RBI) is used to express an increase in the probability of a beneficial outcome (such as returning to work).

The formula that can be used for calculating relative risk reduction or relative benefit increase is:

\[
\frac{(CER - EER)}{CER} \times 100
\]

The CER is simply the proportion of participants in the control group who experienced the event of interest. In our study of exercise to prevent falls, this was 25% (or 0.25). Similarly, the EER is the proportion of participants in the intervention (experimental) group who experienced the event of interest. This was 5% (or 0.05) in the exercise study.

Using our exercise study as an illustration, relative risk reduction can be calculated as: \((0.25 - 0.05) / 0.25 = 0.80\). We could multiply this by 100 to get 80% and report that the participation in an exercise group reduced the risk of falls by 80%.

Alternatively, if you already know the relative risk, then the relative risk reduction can be calculated easily using the formula:

\[RRR = 1 - RR\]

In our example, where the relative risk (RR) is 0.20, the relative risk reduction is 0.80 or 80% (calculated as \(1 - 0.20\)).

The main difficulty in the use of relative risk reduction is that it does not reflect the baseline risk of the event. This means that, as a reader, you are unable to discriminate between small intervention effects and large ones. Baseline risk has an important role to play. To understand this, you need to think about the risk in different populations. For example, in the general population the risk of falling might be between 8% and 40% (depending on factors such as gender, age, comorbidities, and so on). Reducing the risk of falls in a population that has a low risk to start with is very hard to achieve, and any risk reduction that is found would be fairly small (and as you will see in the next section, absolute risk reduction would be smaller still). However, in a population of people who more commonly have falls (such as older adults and those who have previously fallen), the risk of recurrence is much higher and it would, therefore, be easier to achieve a larger relative risk reduction. Interventions that reduce the risk in populations who are at high risk of the event under consideration are likely to be clinically worthwhile.

Another difficulty with the use of the relative risk reduction concept is that it can make results seem more impressive than they really are. Expressing the effects of an intervention in relative terms will result in larger percentages than when the same intervention is expressed in absolute terms. For example, suppose the use of a particular type of mattress reduced the risk of pressure sores from 0.05 to 0.025. In relative terms, the mattress reduces the risk by 50% (calculated by: \(0.025 / 0.05 = 0.5\), or 50%), while in absolute terms it reduces the risk by 2.5% (calculated by: \(0.05 - 0.025 = 0.025\), or 2.5%). Similarly, in the hypothetical trial on exercise for reducing occurrence of falls the relative risk reduction seems an impressive 80% but, as we will see later, the absolute difference in the rate of events between groups is only 20%. So, we can see that the concept of relative risk reduction can inflate the appearance of intervention effectiveness.

You should also be aware that two studies might have the same relative risk reduction but there may be a large difference in the absolute risk reduction. As we saw in the example above about reducing the risk of pressure sores by using a particular type of mattress, a reduction in
Evidence about effects of interventions

The risk of a fall from 0.05 to 0.025 reduces the relative risk by 50%, while in absolute terms it reduces the risk by 2.5%. Let us say that another study found that the risk of pressure sores in a very high-risk group of people was reduced from 0.8 to 0.4 because of the intervention (the mattress). In relative terms, although the mattress reduced the relative risk by 50% (calculated by: \(0.8 \div 0.4 = 0.5\), or 50%), the absolute risk reduction is 40%, which is a more clinically valuable result than the absolute risk reduction of 2.5% that the first study found.

**Absolute risk reduction (or absolute benefit increase)**

Another way of presenting information about dichotomous outcomes is by referring to the absolute risk reduction. The absolute risk reduction is simply the absolute arithmetic difference in event rates between the experimental (intervention) and the control groups.41 This is shown graphically in Figure 4.7. Absolute values are simply the value of a number, regardless of its sign (positive or negative sign). The notation used for absolute values is vertical bars either side of the value, for example: |x|. The absolute arithmetic difference for risk reduction is calculated as: \(\text{EER} - \text{CER}\). As with the previous methods of calculating risk that we have explained, there is a flip side to this concept. This is known as absolute benefit increase and is used when referring to a beneficial outcome (such as being discharged home instead of discharged to residential care).

In our hypothetical exercise study for reducing the risk of falling, the absolute risk reduction would be 25% - 5% = 20%. This could be expressed as: ‘the absolute risk of having a fall was 20% less in the people who were in the exercise group compared with those in the control group’.

A big absolute risk reduction indicates that the intervention is very effective, but how big is big enough to be considered clinically significant? A more meaningful measure, known as ‘number needed to treat’, can be used instead.

**Number needed to treat**

We saw that the absolute risk reduction of having a fall in the exercise study was calculated as 20%, but is this clinically worthwhile? Number needed to treat (NNT) is a method of making the magnitude of the absolute risk reduction more explicit and is a more clinically useful concept.42 The number needed to treat is simply the inverse of absolute risk reduction42 and is calculated as:

\[
1 \div (\text{EER} - \text{CER})
\]

It tells you the number of people who would need to be treated to achieve the event of interest once. In the trial of exercise for preventing a fall, the number needed to treat is \(1 \div (0.25 - 0.05) = 5\). So, in this example, you would have to treat 5 people for 6 months with the hypothetical exercise intervention to prevent one (extra) person from having a fall (when compared with the control group). Obviously, a smaller number needed to treat is better than a large one. An intervention that has a smaller
number needed to treat is more effective than an intervention that has a larger number needed to treat.

This concept of number needed to treat makes it easier to consider clinical significance, as you can more easily weigh up the benefits of preventing the event in one person against the costs and harms of providing the intervention. The size of the smallest clinically worthwhile effect (that is, the smallest worthwhile number needed to treat) will then depend on the seriousness of the event and the costs and harms of the intervention.

Another handy feature when using number needed to treat is that you can compare two different interventions that are trying to achieve the same outcome that have the same number needed to treat, but have other features that are different. For example, one of the interventions may have a shorter intervention time and/or result in fewer side effects and/or be more convenient to patients and/or be less expensive. It is important to consider features such as these that a particular intervention has when making a decision about which intervention to use with a patient.

**Applying results to your clinical situation**

Most health professionals find it difficult to translate the results from studies to individual patients or to specific clinical situations, as studies usually only tell us about the average effects of the intervention. Further, the participants who took part in the study may be different from the patients that we see, and the intervention may differ from the intervention that we use. However, as a rough guide, if your patient is ‘healthier’ or the situation that you are considering is more optimistic than that in the study, the number needed to treat would be lower, the relative benefit increase would be higher, the mean difference would be larger, and so on. If, however, your patient is ‘worse off’, or the situation that you are considering is worse than that in the study, the number needed to treat would be higher, the relative benefit increase would be lower, the mean difference would be smaller, and so on. A solution to applying the results of a trial to patients with higher or lower levels of risk has been described by Straus and Sackett and you may wish to read their article to learn more about this.

**Dichotomous outcomes—precision of the treatment effect**

Confidence intervals are also important to consider when examining dichotomous outcomes as, again, they indicate how much uncertainty is associated with the estimate of the intervention effect (in other words, the precision or accuracy of the estimate). The principles of confidence intervals associated with dichotomous data are similar to those for continuous data, but there is one very important difference to consider—what is the appropriate ‘no effect’ value to use?

- For effect size estimates where subtraction is involved, such as mean differences (for continuous outcomes) and absolute risk reductions, the ‘no effect’ value is 0.
- For effect sizes that involve division, such as risk ratios and odds ratios, the ‘no effect’ value is 1.

Therefore, you can see that it is important to consider the type of effect measure you are interpreting in order to reliably interpret the confidence interval. The same general principle applies, though—a 95% confidence interval that does not include the ‘no effect’ value indicates that the result is statistically significant. Table 4.1 presents a summary of the dichotomous effect measures, including what their ‘no effect’ value is, that have been discussed in this section of the chapter.

As we could do with continuous outcomes, if the confidence interval has not been provided by the authors of the article it is sometimes possible to calculate an approximate confidence interval. This is illustrated in Box 4.3. Again, a simplified version of the more complex equation is used, but the confidence interval that is calculated is sufficient for use by health professionals who wish to
<table>
<thead>
<tr>
<th>Type of measure</th>
<th>Definition</th>
<th>Formula</th>
<th>‘No effect’ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative risk (RR) (also called ‘risk ratio’)</td>
<td>The ratio of the probability of the event occurring in the intervention group versus the control group. Expressed as either a decimal proportion or a percentage.</td>
<td>The risk of an event in the intervention group divided by the risk in the control group: EER ÷ CER</td>
<td>1</td>
</tr>
<tr>
<td>Relative risk reduction (RRR)</td>
<td>The proportion of the risk that is removed by the intervention. That is, the proportional reduction in an event of interest in the intervention group compared with the control group. Usually expressed as a percentage.</td>
<td>(CER – EER) ÷ CER × 100 or 1 − RR</td>
<td>0</td>
</tr>
<tr>
<td>Absolute risk reduction (ARR)</td>
<td>The absolute arithmetic difference in event rates between the experimental (intervention) and control groups. Usually expressed as a percentage.</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Number needed to treat (NNT)</td>
<td>The number of people that would need to be treated for the event of interest to occur in 1 person.</td>
<td>1 ÷ ARR or 1 ÷ (EER – CER)</td>
<td>Infinity.</td>
</tr>
<tr>
<td>Odds ratio (OR)</td>
<td>The odds of an event in the intervention group divided by the odds of an event in the control group. Usually expressed as a decimal proportion.</td>
<td>[EER/(1 − EER)] ÷ [CER/(1 − CER)]</td>
<td>1</td>
</tr>
</tbody>
</table>

CER = Control Event Rate; EER = Experimental Event Rate.

**How to calculate the confidence interval (CI) for absolute risk reduction**

A formula⁴⁰ that can be used is:

\[
95\% \text{ CI} = \text{Difference in risk} \pm 1/\sqrt{n_{av}}
\]

where \(n_{av}\) = average of the group sizes.

For our study of exercise, the 95% confidence interval for the absolute risk reduction would be:

\[
= (30\% − 5\%) \pm 1/\sqrt{100} = 25\% \pm 0.1 = 25\% \pm 10\%
\]

This tells us that the best estimate of the absolute risk reduction achieved from participating in an exercise group is 35% and that the 95% confidence interval extends from 15% to 35%.

use the information to assist in clinical decision making. Once you know the confidence interval for the absolute risk reduction, if you wish you could plot it, the effect estimate and the smallest clinically worthwhile effect on a tree plot in the same way that we did for continuous outcomes in Figure 4.5.

Calculating the confidence intervals for number needed to treat is fairly straightforward, as you just use the inverse of the numbers in the confidence interval of the absolute risk reduction. However, understanding how to interpret the confidence intervals for number needed to treat can be complicated, and the article by Altman is recommended to understand this in detail.

**CLINICAL SCENARIO (CONTINUED)**

**MAIN RESULTS—DICHOTOMOUS**

In the CBT for depression for people with heart failure trial, the authors dichotomised the scores from the BDI-II so that participants with scores of \( \leq 9 \) were considered to have achieved a remission from depression, while those scoring above this did not. Note that in using the data this way, achieving a remission is a *positive* change and therefore we need to think in terms of *benefit* rather than *risk*. The authors indicate that 46% of participants in the CBT group achieved remission from depression, compared with 19% in the control group. From this we can work out that the **absolute benefit increase** is \( 0.46 - 0.19 = 0.27 \) (27%). The **number needed to treat** is the inverse of the absolute benefit increase, so \( 1/0.27 = 3.70 \). The authors provide the 95% confidence interval of the number needed to treat as 3.62 to 3.90. Rounding the number needed to treat up to 4 (since you can’t have 3.76 people) indicates that you would have to treat 4 people for 6 months with CBT for one extra person to achieve a remission from depression. With this information, you can, along with your patient, weigh up the likelihood of achieving remission from depression against the costs, risks and inconveniences of the intervention.

When making sense of the results of a randomised trial, one further issue that you should be aware of is that trials usually report results from many different outcomes. With so much information to process it may be helpful, in some cases, for you to focus your attention on the main outcome(s) of the trial, the outcome(s) of interest to the question that you initially formed and/or the outcome that is of interest to your patient. If you have got to this point and determined that the article that you have been appraising not only appears to be reasonably free of bias but also contains some clinically important results, you then proceed to look at whether you can apply the results from the study to your patient or clinical situation. This is the third and final step of the critical appraisal process that we described in Chapter 1.

**HOW CAN WE USE THIS EVIDENCE TO INFORM PRACTICE?**

If you have decided that the validity of the study you are appraising is adequate and that the results are clinically important, the final step of the critical appraisal process is to consider the application of this information to the clinical situation that prompted your original clinical question. To do this, there are a few important questions that should be considered:

- Do the results apply to your patient or situation?
- Do the benefits found outweigh any harm, costs and/or inconveniences that are involved with the intervention?
- What other factors might need to be considered when applying this evidence?
Do the results apply to your patient or situation?

When considering whether the results apply to your patient or situation, you essentially need to assess whether your patient is so different from the participants in the study that you cannot apply the results of the study that you have been reading to your patient or situation. So, rather than just thinking about the study’s eligibility criteria with respect to your patient, are the differences problematic enough that the results should not be applied? Further to this, results can sometimes be individualised to a particular patient by considering the individual benefit and harm for that patient. There is some complexity involved in doing so, which is beyond the scope of this book. If you wish to learn about this further, you are advised to read the article by Glasziou and Irwig45 to understand how this may be done.

Do the benefits found outweigh any harm, costs and/or inconveniences that are involved with the intervention?

Understanding whether benefits outweigh harms, inconveniences or costs requires using information about benefits that is given in terms of the size of the intervention effect (for example, mean differences, relative risk reduction, number needed to treat) and comparing this against possible harms or even inconveniences. When information about harm and cost is provided in the article, this is relatively straightforward. However, this type of information is often not provided and some estimate of the costs associated with the intervention might need to be made. Information about harm is obviously more difficult to estimate, and other sources might need to be consulted to get a sense of the potential harms involved. This is a grey area of practice in which clinical experience and discussion with patients about their preferences and values become very important.

What other factors might need to be considered when applying this evidence?

A further important question to consider is: what other factors might need to be considered when applying this evidence? When you are thinking about what other factors affect the delivery of an intervention, there are a few key questions that you can ask, such as:

- How much does it cost?
- How many sessions might be involved or how long would the patient need to stay in hospital?
- How far would the patient (or the health professional) need to travel?
- Are the resources (for example, any equipment that is needed) available to deliver the intervention?
- Do you or other health professionals working with you have the skills to provide the intervention?
- Are the details of the intervention sufficiently described in the article or available elsewhere so that intervention can be accurately provided?

A central component of applying research evidence to your patient is discussing the information with them. The success of many interventions is dependent on the health professional successfully providing the patient with appropriate information.

Involving the patient in the decision-making process is important, and this is discussed further in Chapter 14. In Chapter 1 we emphasised the need for integrating information from patients, clinical experience, research and the practice context. To do so is the art of evidence-based practice. Integrating so many pieces of information from many different sources is certainly an art form and one that requires clinical reasoning and judgment. The roles of clinical reasoning and judgment are discussed further in Chapter 15.
CLINICAL SCENARIO (CONTINUED)

USING THE EVIDENCE TO INFORM PRACTICE

So, what can we conclude from our reading of the CBT trial article? You know that depression is common among people with heart failure, that it is associated with an increased risk of mortality and poorer quality of life, and that tackling depression is an important issue. In summary, this trial was a well-constructed trial, although it was unable to achieve blinding of participants, health professionals or study personnel. This leads you, as the reader, to be a bit cautious of the results reported. For example, the depression data were collected by participant self-report, which would be subject not only to error but also to differential bias (that is, there could be systematic differences in the self-reports of people in the intervention compared with those in the control group due to their knowledge about which group they were in). Participants in the intervention group also received more contact over the time of the trial, which makes you wonder what part that extra attention might have played in influencing outcomes.

At 6 months, the intervention group had lower levels of depressive symptoms than the control group (a mean between-group difference of –4.43) and this was sustained at the 9- and 12-month follow-ups. However, even though the 6-month result was statistically significant, you consider its clinical significance to be uncertain and you have similar concerns about the effect on anxiety at 6 months. Rates of remission from depression were higher in the intervention group than the control group, with an NNT of 4. As the reader of this article, you need to consider how the biases present in this study might have affected these results. You then also need to consider how to apply these results in your situation.

Do the results apply to our patient or situation?
The group of people that we were concerned about were people with heart failure who also had depression. The study that we looked at involved men and women with a mean age of 55.8 years who had a depression score of ≥14 on the BDI-II. It is likely that these results do apply to the group of people that we are concerned about.

Are the benefits worth the harms and costs?
While there are few ‘harms’ associated with this intervention, CBT can explore painful feelings, emotions and experiences that may, at times, make people feel emotionally uncomfortable. No details about costs of the intervention are provided in the article. All we know is that the intervention involved 1 hour of CBT per week and tapered off to monthly treatment towards the end of the 6 months, with a number needed to treat of 4. A very general idea of costs can be determined from this information. For individual patients, the ‘cost’ to them can be thought about in terms of the time commitment, inconvenience, potential disruption of other activities, and travel involved in attending weekly CBT sessions.

Other factors to consider
Other practical factors to consider are the availability of this type of CBT program in your clinical vicinity or, if you were to run it at the community health centre, the resources (such as staffing and space, or any training or intervention materials needed) and costs that would be involved. Data about the cost-effectiveness of this intervention would be valuable to further facilitate your team’s decision making about whether to implement this intervention or not.
**SUMMARY POINTS OF THIS CHAPTER**

- The effects of intervention are best determined through rigorous randomised controlled trials (or, better still, systematic reviews of randomised trials), as their methods best reduce the risk of bias.
- Most randomised controlled trials are not free from bias. Key questions that should be asked when appraising the risk of bias in a randomised trial are:
  - How were participants randomised to groups?
  - Was the allocation sequence concealed?
  - Were the groups similar at baseline?
  - Were participants, health professionals and study personnel blind to group allocation?
  - Was follow-up adequate?
  - Was intention-to-treat analysis used?
- Our understanding of the risk of bias that is present in a study is affected not just by the methods that the researchers used, but also by how well these are reported in the article that describes the study.
- Two factors to consider when making sense of the results of a randomised trial are:
  - the size or magnitude of the intervention effect (this may be provided as continuous data or dichotomous data); and
  - the precision of the intervention effect (which can best be determined by inspecting the confidence interval for the intervention effect).
- Applying the results of the study requires thinking through whether the results apply to your patient or situation, a range of logistical factors that can affect the delivery of the intervention, and, in collaboration with your patient, discussing whether the possible benefits outweigh any harm, costs or inconveniences.

**References**

5. Guyatt GH, Sackett DL, Cook DJ. Users’ guides to the medical literature. II. How to use an article about therapy or prevention. JAMA 1993;270:2598–601.