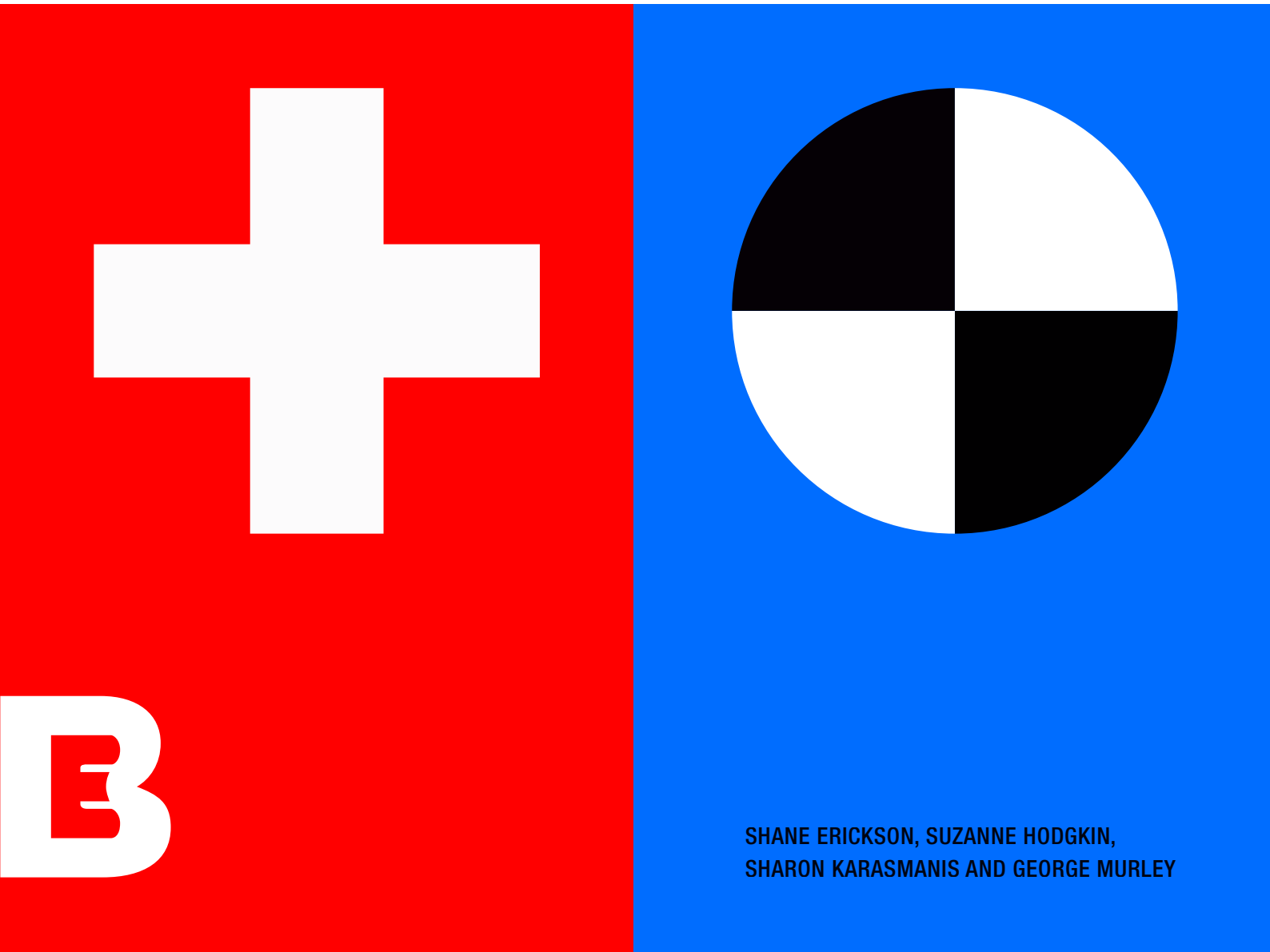
RESEARCH AND EVIDENCE IN PRACTICE



SHANE ERICKSON, SUZANNE HODGKIN,
SHARON KARASMANIS AND GEORGE MURLEY

LA TROBE EBUREAU
La Trobe University, Melbourne, VIC 3086, Australia
https://library.latrobe.edu.au/ebureau/

Published in Australia by La Trobe University

First published 2018

Research and Evidence in Practice
Shane Erickson, Suzanne Hodgkin,
Sharon Karasmanis and George Murley
ISBN: 978-0-9953727-7-1
DOI: https://doi.org/10.26826/1007

Other information
Designed by La Trobe eBureau
Edited by Hilary Cadman
Cover designed by Evi O. Studio
This etextbook has been peer reviewed.

Enquiries: library@latrobe.edu.au

Contents

Authors

SHANE ERICKSON

Photo by La Trobe University Photography Department; used with permission

Dr Shane Erickson is a lecturer, researcher and speech pathologist at La Trobe University in Melbourne, Australia. Shane has taught evidence based practice (EBP) and research methodology subjects to students across the continuum of undergraduate and masters level courses. He is passionate about helping students understand the important role that research evidence plays in guiding clinical decision-making. This is certainly something that he is acutely aware of in his own clinical practice, which consists of treating people who stutter across the lifespan. Shane recognises that for many students the skills required to find, appraise and apply research into practice don't come easily, and so teaching them in a way that is accessible and engaging for all students is critical. Shane has published a considerable number of research papers related to the use of technology in the delivery of speech pathology interventions, research translation and teaching and learning, and has also recently presented papers at several international conferences.

SUZANNE HODGKIN

Photo by Isabelle Hodgkin; used with permission

Dr Suzanne Hodgkin is an Associate Professor and Deputy Director of the John Richards Centre for Rural Ageing Research (JRC). Prior to working as an academic, she had considerable practice and management experience in the government and non-government sectors. Her masters' research explored the problem of recruitment and retention of frontline rural child protection staff in Victoria. Suzanne Hodgkin graduated with a PhD in social work and social policy in 2006. She has extensive teaching experience having taught in the Rural Health in clinical skills, human rights and ethics and in research methods.

A/Professor Hodgkin is an internationally recognised researcher. Her principal research interests lie in the area of Health and Aged Care Workforce, Models of Aged Care and Social Capital. The common theme is the provision of services and supports to older people living in rural communities. Suzanne, considered an expert in a variety of research methods including mixed methods research, has had her research design work published in the Journal of Mixed Method Research and in mixed methods research texts. She has published her work in several international and national journals. She has also disseminated findings from her research to international and national conferences.

SHARON KARASMANIS

Photo by La Trobe University Photography Department; used with permission

Sharon Karasmanis is the Manager of Learning and Teaching at La Trobe University Library, and is responsible for curricular and co-curricular programs and services across six campus libraries. Sharon's research interests have included student information seeking behaviour, and research skills for evidence-based practice in Health Sciences; as well as recent research into students' digital experience in academic settings. From this research, emerged a sound pedagogical model for embedding information and digital literacies into the curriculum. In 2013, Sharon collaborated with the Faculty of Health Sciences on the development of the award winning Research and Evidence modules, designed as an introduction to evidence-based practice teaching to first year Health Sciences' students. Sharon was awarded two Trobe University Citations for Outstanding Contribution to Student Learning in 2013 (team citations), and in 2014 with a Library team, an OLT Award for Outstanding Contribution to Student Learning, from the Office of Learning and Learning in Australia.

GEORGE MURLEY

Photo by George Murley; used with permission

Dr George Murley is a podiatrist with a research background in the field of lower limb biomechanics. His PhD research investigated biomechanical differences between those normal and flat-arched feet and the influence of foot orthoses. In the field of biomechanics, his research is specialised in the area of electromyography (electrical activity of muscles during movement). Dr Murley has more than 30 peer reviewed journal publications and a book chapter titled 'Effects of footwear on muscle function'. In 2010, he received the Stephen Duckett research award for the best doctoral thesis in the Faculty of Health Sciences at La Trobe University. Dr Murley consults as a full time podiatrist in private practice and is an Honorary Associate at La Trobe University.

Introduction

This etextbook is an introduction to the use of research-based evidence in professional health-care practice. It will help you to understand the role of evidence in health care, including how evidence is developed, and how to interpret research methods and outcomes. This etextbook explains how to use systematic search methods to obtain, interpret and summarise key design elements of peer-reviewed journal articles or other forms of evidence-based material. Finally, it outlines how to identify, discuss and interpret selected research outcomes and basic statistics from peer-reviewed journal articles or other forms of evidence-based material – and to estimate the relevance and importance of these outcomes to clients. Throughout this etextbook, we will refer to patients and consumers as clients. The etextbook is written in an informal style with a mix of text, illustrations and embedded videos. We hope that it will help you to develop a sense of understanding and mindfulness in executing evidence-based practice (EBP).

Think of the last time you visited a doctor, perhaps because of a cold or flu, or something more serious. Whatever your condition, during the consultation your doctor would have been quickly processing a lot of information. He or she would have been analysing your symptoms via an assessment, establishing the appropriate intervention, counselling you on steps to take to minimise any pain or discomfort, and then deciding whether you needed to have time away from study or work. Each of these aspects of your care required a decision, and each decision required a rationale. Your doctor would have been aware that an incorrect or unfounded decision about any aspect of your care might have meant, in the case of a cold or flu, that you spent an extra few days with that annoying runny nose and cough. However, in the case of a more serious health condition, an incorrect or unfounded decision could have had potentially dire consequences.

EBP has its origins in evidence-based medicine, but has now become commonplace in all areas of health practice. It requires professionals to make decisions about practice that are supported by the best available evidence, coupled with professional expertise and available resources; these decisions must also take into account the rights, values and preferences of patients, clients and consumers. This holds true, whatever the nature of the decision. So, for example, an evidence-based decision could concern the best treatment, therapy or diagnostic procedure for an individual patient or client; the best way to implement a community health promotion program; or the best way to prevent the spread of a particular disease in a population.

EBP promotes an attitude of inquiry and should lead you to ask questions such as: Why am I doing things in this way? Could I be more effective? Could I make my service more targeted? In asking these types of questions, health professionals become more accountable. We expect the health professionals

we visit to have our best interests in mind. Once you start to practice, you will also have people depending on your decisions and advice. As part of providing a professional service, you have a responsibility to ensure that, whenever possible, you use the best available evidence to inform your practice.

Interactivity

This etextbook has multiple interactive elements that extend and expand upon the content. It is recommended that Adobe Acrobat be used to utilise these interactive elements.

VIDEOS

Videos are embedded throughout the etextbook. These videos are presented by specialist academics to engage and further develop the discussion; click the play button on the video to watch.

ACTIVITIES

Fillable tables allow the reader to complete activities and engage with the content in a different way. Look out for practice-related questions and scenarios that require readers to complete a table with information.

GLOSSARY RESEARCH TERMS

Look out for Glossary Research Terms throughout the text; when clicked these terms link to the glossary at the end of the etextbook. (Use the command *alt + left arrow* to return to your place in the PDF).

ANNOTATION

Annotating this etextbook is a useful tool and is great for taking notes on the pages. The easiest way to add comments is to use the sticky note tool in Adobe Acrobat. For more information and instructions on how to annotate PDF's there is a tutorial on the Adobe website.

To cite this resource

Erickson, S., Hodgkin, S., Karasmanis, S. & Murley, G.S. (Eds.), (2018). *Research and evidence in practice* [E-Resource]. Retrieved from https://doi.org/10.26826/1007.

Chapter 1

Evidence for practice: what, why and how?

Introduction and learning outcomes

This section gives an overview of the key concepts that underpin the integration of research evidence into health-care practice. The term Evidence Based Practice is used frequently throughout this etextbook, but what do we mean by EBP, and why will it be important for you in your professional career?

In the following video Simon Pampena gives an overview of research and evidence for health sciences:

https://doi.org/10.26181/5c11903941770

KEY LEARNING OUTCOME

Describe the background, rationale, key principles and processes of EBP.

ENABLING OUTCOMES

Once you have completed this chapter, you should be able to:

- identify why EBP is important
- define EBP and essential related concepts, principles and processes (including the five-step EBP model)
- explain key differences between two broad research approaches (i.e. qualitative and quantitative research)
- identify factors that affect the quality of evidence and research (e.g. **validity**, **reliability** and generalisability)
- distinguish different types of evidence (e.g. **primary research** and **secondary research**, reviews and guidelines)

What is evidence?

Let us consider what is meant by the term 'evidence' in the context of EBP. The doctor in the example on page VII, is likely to have used research evidence as a rationale for the decisions made about your care. Research evidence – that is, evidence generated by studies that use systematic processes to address questions about a specific aspect of health – is the main type of evidence used in EBP in health. Such evidence comes, for example, from studies that investigate the effectiveness of a treatment or a preventative measure, the cause of a specific disease or health condition, and the experiences of people who are living with a particular disease or health condition. You will learn more about specific types of research in subsequent chapters.

Although you may not have read any of the types of research evidence mentioned above, you should be broadly familiar with the term 'evidence'. In fact, you have probably searched for evidence before. Think back to a time when you have made an important purchase – perhaps a new car or mobile phone. Such a purchase is a big decision and takes a lot of consideration. Very few people just turn up to a car yard and purchase the first car they see! In the case of a new car, most people have some specific criteria in mind before they decide on the most appropriate make and model to buy. One of these criteria might be **reliability**, given that no-one wants to be landed with big repair bills. To determine how reliable a particular type of car is, you might seek information from different sources, such as your parents, friends, websites, car owner's forums and car salespeople. Each of these sources could provide you with information about the reliability of different makes and models of cars, and thus influence your purchase.

Of course, the tricky part is working out which information to believe and whether you can trust the source. Indeed, as you will read later in this etextbook, the process of searching for evidence in an EBP context shares many similarities with a search for evidence for an important purchase – in both cases, you have to work out whether the evidence is trustworthy.

The research that underpins evidence-based decisions comes in many shapes and sizes, and is often more than just a simple, pure experiment. For example, research can involve:

- monitoring the incidence and prevalence of specific diseases and health conditions, as Australia does with its regular National Health Survey
- analysing data collected in the course of monitoring specific health interventions; for example, monitoring the number of adverse events that occur in conjunction with a vaccination program for seasonal influenza
- producing evaluation reports of a health intervention; for example, producing reports on a specific health promotion program

Principles of evidence-based practice

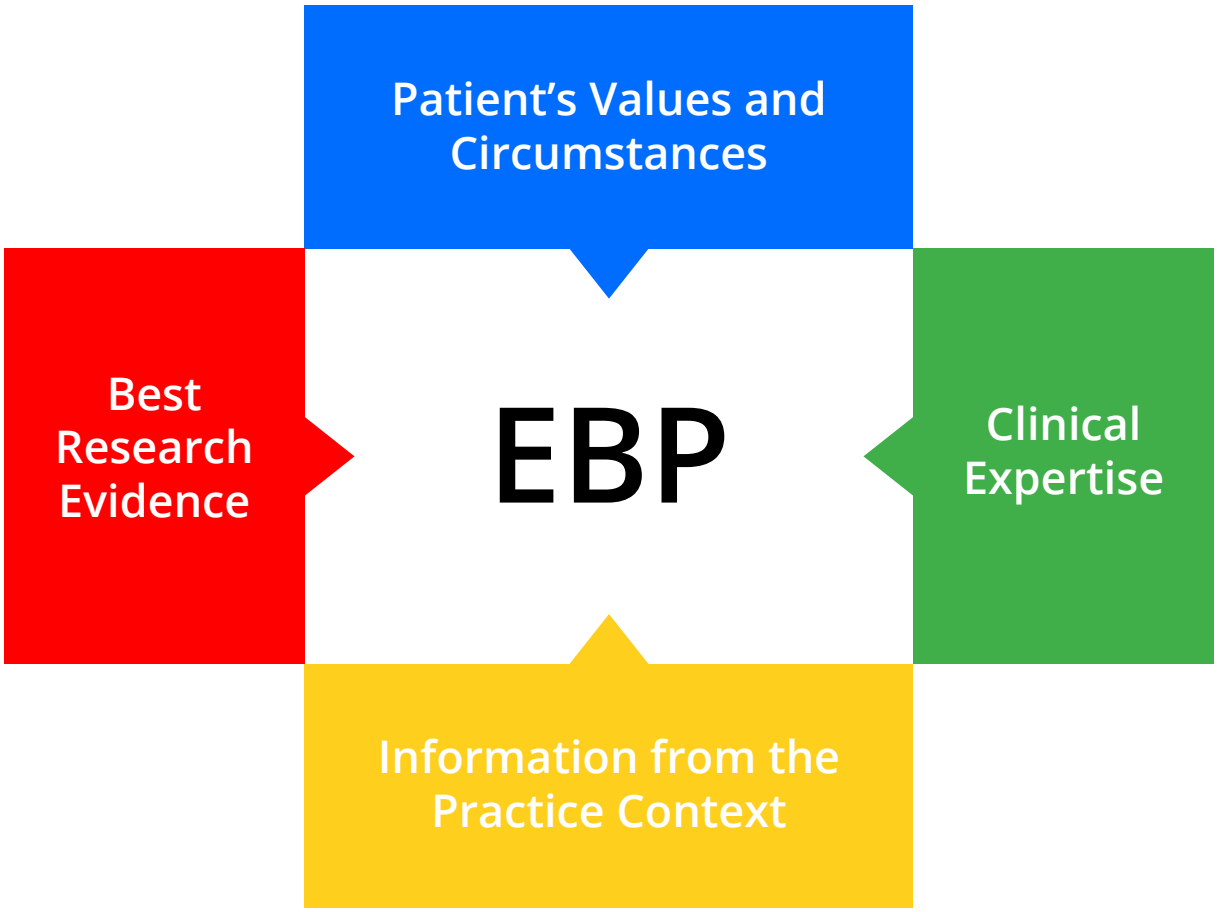
Given the importance of the decisions you will make as a health practitioner throughout your career, it is important for you to be familiar with the two key principles of EBP.

The ***first*** key principle of EBP is that decisions about practice should be supported by the following:

> **Best research evidence** – that is, valid and clinically relevant research (discussed below) that has been conducted either internally (i.e. within your professional environment) or externally (i.e. by other researchers). Both forms of research should be considered in the decision-making process.
>
> **Professional and clinical expertise** – that is, the skills and past experience that help to identify each patient's health state and diagnosis, and the risks and benefits of potential interventions.
>
> **Information from the practice context** – that is, the clinical circumstances and setting in which you are working, and the availability of resources, space and time that would be needed to help you implement a specific intervention. Your decisions also need to take into account the patient's comorbidities (i.e. other conditions).
>
> **Client's values and circumstances** – that is, the rights, values and preferences of patients, clients and consumers. To serve the patient, the unique preferences, concerns and expectations of each patient must be integrated into clinical decisions.

Figure 1.1 demonstrates how each of these elements is equally weighted within the EBP framework. This is important because EBP is not intended to be a one-size-fits-all solution, where professionals blindly adhere to the findings of research studies alone. After all, even recommendations based on excellent research evidence may be inappropriate for a given client's or patient's unique situation.

Figure 1.1 An EBP framework gives equal weight to research evidence, clinical expertise, information from a practice context and client's values and circumstances.



Adapted from diagram in 'Evidence-based practice in health' by Murray Turner from University of Canberra used under CC BY-SA 4.0

The *second* key principle is, that to maintain its currency, EBP requires ongoing professional development.

Thus, it requires that health practitioners are obliged to maintain their currency of practice in a system in which practice must change and respond to new knowledge. The skills and knowledge you will learn from this etextbook will serve you well into the future. Given the changing nature of health care, the things you learn about today may be outdated in a matter of years. However, if you have the skills to be constantly updating your knowledge base, then you are future-proofing yourself as a professional.

Implementing evidence-based practice

There are several models of EBP that effectively formalise the processes involved. One such model is a five-step approach, which can be summarised as follows:

Step 1: Ask an answerable practice-related question.

Step 2: Acquire relevant evidence to answer the practice-related question.

Step 3: Appraise the acquired evidence.

Step 4: Apply the appraised evidence to practice.

Step 5: Assess your own performance in executing Steps 1–4, and set learning goals to improve your future performance.

Step 1: Ask an answerable practice-related question

The process begins with you recognising that you need some information, whether it be about an intervention, a diagnosis, the **aetiology** of a health condition or the patient experience of living with a health condition. An important step is turning this need for information into an answerable question to be investigated. This process is explored in Chapter 2.

Step 2: Acquire relevant evidence to answer the practice-related question

To acquire the relevant evidence, you need to conduct a search. This is usually done by searching databases that index articles that report research studies and systematic reviews that are relevant to the question. This process is explored in Chapter 3.

Step 3: Appraise the acquired evidence

Having found evidence by searching one or more databases, you now need to critically appraise it according to certain criteria. Essentially, you need to work out whether the evidence is worthy of being used to inform your decision-making. For various reasons, not all published research is of good quality. Chapter 4 discusses some of the things that can affect the quality of a study and thus potentially provide misleading results. You need to be able to recognise whether a study was conducted in a way that means you can trust the results.

Step 4: Apply the appraised evidence to practice

Once you have appraised the evidence and decided that it is worthy of being used, you then need to apply it to the aspect of practice that was the subject

of your initial question. For example, if you have found strong evidence that a new intervention is highly effective and efficient, you may decide to use it in practice. This is the point at which you integrate your expertise, your patient's needs and the context you are working in. Remember that any decision must take into account the unique needs, values, preferences, concerns and experiences of your particular client or patient.

Step 5: Assess your own performance in executing Steps 1–4, and set learning goals to improve your future performance

The final step is to evaluate your performance on all of the previous steps, with the aim of improving your future performance in applying the EBP process. You need to be able to complete this process efficiently and effectively, so that it does not become a time-consuming task. By asking yourself questions that promote self-reflection, you can reflect on what you are doing well and what you could do better.

More detail will be provided on these 5 steps in later chapters.

Research evidence

Much of the evidence that is drawn upon for evidence-based decision-making is generated by research studies. A research study and its findings are usually reported in journal articles (i.e. articles in scholarly journals). Usually, journal articles are peer reviewed before they are published; that is, they are critically appraised by people who have expertise in the area with which the study is concerned. However, some journals rely solely on editorial review.

Research involves a process of systematic investigation, which often involves the following steps:

- formulating a research question
- conducting a review of the relevant literature
- deciding on a method for addressing the research question
- using the chosen method to collect data
- analysing the data
- interpreting the data in terms of the answer to the research question

This list is by no means exhaustive. Additionally, the nature of research means that this is often a cyclical process, and answering a research question often leads to more questions, and the need to repeat this process.

For example, consider a research study to investigate a vaccine for the H1N1 influenza virus. The process would start with a search for literature relevant to the issue; this would include information about the incidence and prevalence of influenza caused by the H1N1 virus, any previous studies of H1N1 vaccines, and any studies of vaccines for other types of influenza.

This previous research would form part of the basis for the research question. Research questions need to be both specific and worded in such a way that they are directly answerable. Consider these two examples of research questions that could be asked for this study:

> How useful are H1N1 vaccines?
>
> How effective is a (specific or new) vaccine for the H1N1 virus, in people aged 65–85 years living in Australia?

The second question is preferable because it is more specific. In response to this second question, a trial might be designed involving a sample of people aged 65–85 years, with half of the people randomly selected to receive the new vaccine and the other half to receive the current vaccine. The data would then be analysed using statistical procedures and interpreted in terms of the answer to the study question; that is, in terms of how effective the vaccine was found to be for people aged 65–85 years living in Australia.

After publication, reports of many research studies may be incorporated into reviews (i.e. papers that summarise other papers). Studies in which the authors collect original primary data, such as the one described above for influenza vaccines, are referred to as '**primary research**', whereas papers that review primary research are referred to as '**secondary research**'. One type of review is a systematic review, which starts with a comprehensive search for reports of studies that addressed a specific research question. The studies are then vetted to determine whether they meet a set of criteria for inclusion in the review. Systematic reviews, and journal articles reporting single studies, constitute the bulk of the evidence that is used in evidence-based health practice.

8

Types of primary research studies

Several different study types are commonly used in health research. The broadest distinction between study types is in terms of whether they are *quantitative* or *qualitative*. Some research combines these types of studies, and is referred to as *mixed methods* research. These broad study types are outlined below and are discussed in more detail in Chapters 5-11 (quantitative

research), Chapters 12-13 (**qualitative research**) and Chapter 14 (mixed methods research). The most appropriate type of study will depend on the particular purpose of the research.

QUANTITATIVE

Quantitative research seeks to test theories or hypotheses by analysing relationships. Often, these theories initially arise from an observation. For example, imagine that you notice that your cat watches TV when birds are on it, but never when any other animal or program is on. You decide to investigate this, and you ask 20 cat owners to keep diaries about their cat's TV viewing habits. When you analyse these diaries, you find that they show that many cats watch TV more when there are birds on the screen. Your analysis might include a measurement of the amount of time each cat spent watching TV when a bird was on the screen, compared with a measurement of the amount of time each cat watched TV when there was no bird on the screen. Analysing this relationship could lead you to deduce that cats really do watch more TV when birds are on the screen!

'Fatty watching himself on TV ' from Wikimedia Commons used under CC BY 2.0

Quantitative studies involve measurement of study participants' characteristics that are relevant to the research question. *Measurement,* in the context of health and well-being, can be defined as the process of quantifying health and related **phenomena**, and assigning numbers to represent characteristics of people and their environments. This process of quantifying characteristics (e.g. physical, behavioural, psychological or social qualities) provides the foundation for all quantitative health research **methods**. In chapters 7-11 you will learn more about measurement processes.

QUALITATIVE

In direct contrast to quantitative research, qualitative studies use a research approach that uses inductive rather than deductive logic for both data collection and data analysis. Its philosophy of understanding phenomena from an insider's point of view sets it apart from quantitative research. It generally answers the questions 'why' and 'how' rather than 'how much'.

Sometimes, **qualitative research** can lead to the generation of theories about health-related phenomena, which in turn can lead to hypotheses that are tested using a quantitative approach.

Qualitative research can make an important contribution to the appropriateness of health practice – in particular, to the provision of appropriate health care. Health practitioners need to have a good understanding of what it means to be ill or to be injured, and what it means to live with an illness or other health condition, or a disability. Understanding the lived experience of illness or disability is particularly important in regard to chronic illnesses and other chronic conditions. Qualitative research is often a better option than quantitative research for providing in-depth understanding of these phenomena. By drawing on evidence from qualitative studies, health practitioners can better understand the experiences – and the associated attitudes, feelings, perspectives and beliefs – of patients or clients. Practitioners will then be able to provide more sensitive and appropriate care for their own patients or clients. In addition, when evidence derived using a qualitative approach is combined with evidence from quantitative studies, it can help practitioners to understand an issue in greater depth, and to appropriately apply the body of evidence in their practice.

Types of secondary research studies

There are a number of different types of review studies, three common types are discussed below:

SYSTEMATIC REVIEWS

A systematic review must conform to strict guidelines; this means that readers can have confidence in the findings from this type of **secondary research**. A systematic review needs to provide explicit details of the steps in the review process, including the strategy used for searching for the evidence, the criteria for including studies in the review and (if applicable) the criteria for excluding studies from the review.

More specifically, the steps in the review process include:

- formulating a question to be answered by the review
- conducting a search of the literature
- using predetermined criteria for deciding which studies should be included in the review

- critically appraising the methodological quality of the individual studies
- extracting the relevant data from each study
- synthesising the extracted data using statistics (where appropriate)
- summarising the overall results of the review and discussing the implications of these results

Of course, a systematic review is only valuable for informing practice if the available evidence for the review comes from high-quality studies. Also, in many areas of health research, there are too few papers to be synthesised into a systematic review. More information on systematic reviews is available here.

NARRATIVE REVIEWS

Not all literature reviews are conducted in a systematic fashion; an alternative is a *narrative* approach, in which the review is conducted in a story-telling fashion. There are basically two kinds of **narrative review**:

- those in which the review constitutes the entirety of the paper
- those in which the review forms only part of the paper, and is integrated into the introductory section of a report on a research study

The quality of narrative reviews varies, given that the analysis, critique and synthesis of the material draws on the creativity and intellectual style of the author. Also, depending on how individuals approach the evidence to be reviewed, they may emphasise different aspects of the evidence, resulting in a potentially different overall conclusion. Taking this to an extreme, some authors may allow their biases to influence the review to the extent that the meaning of the evidence could be completely distorted; for example, a medical researcher employed by a tobacco company may review the literature relating to the health risks of tobacco smoking very differently to an impartial reviewer. A major advantage of systematic reviews over narrative reviews is that the transparent and systematic methodology of a systematic review helps to control for author **bias** and other sources of bias.

CLINICAL GUIDELINES

In many areas of health practice, information from various sources has been compiled into guidelines, to guide health professionals in how to deal with specified clinical conditions. These guidelines can range from simple protocols to high-quality guidelines. The latter are rigorously compiled using a comprehensive review of the research evidence about a particular area, and are often combined with client input and expert opinion. An example is

the Stroke Foundation's Clinical guidelines for stroke management 2017, which may be used by professionals such as occupational therapists, physiotherapists, speech pathologists, nurses and dietitians.

Clinical guidelines are a useful tool because, much like a systematic review, they provide health professionals with an efficient and effective way to digest large amounts of research in specific areas. However, the guidelines add value because they translate this research into recommendations for practice, and thus help health professionals to make better decisions about their clients' care.

Factors that affect the quality of research

A final point about evidence is that all evidence needs to be evaluated for its quality. Before you can use research findings to inform your practice, you need to be sure that you can trust the research you read and the way it was conducted. The degree of certainty about the conclusions that can be drawn from a study's results will vary. As critical consumers of research, you need to be able to recognise key aspects of the way a study is conducted, to determine whether you believe the findings.

In quantitative research, the degree of certainty that we can have about the correctness of conclusions drawn from a study's findings is referred to as 'internal validity'; in **qualitative research**, the trustworthiness of conclusions is referred to as 'methodological rigour'. The notion of internal validity is explored further in Chapter 6, and rigour in Chapter 13, but essentially these concepts relate to the way in which the study was conducted, and the steps the researchers took to demonstrate that the study's findings are believable and trustworthy.

A common threat to the quality of quantitative research is the **validity** and **reliability** of measurements. *Validity* relates to a measurement quantifying

what the study is intended to measure. Reliability relates to a measurement providing the same result every time it is performed. For example, if people in a research trial must have their temperature measured, then the thermometer used should be one that provides consistent results; that is, it provides the same temperature if repeated in the same context. Equally, the choice of thermometer needs to be valid and consistent; for example, the researchers need to decide whether to use a forehead sticker thermometer or an oral thermometer, by considering which is most accurate. As with internal validity and rigour, measurement reliability and validity is further explored in Chapter 6.

Internal validity can be contrasted with 'external validity', which relates to the degree in which the study's findings can be generalised beyond just those people involved in the study. After all, most research is intended to guide practice, and if the results cannot be generalised beyond the small group of people in the study, then the real value of the findings is limited. If the group of people in the study is both large in number and representative of the population, then generally this is said to afford a high level of external validity. Conversely, a sample that is both small and unrepresentative constitutes a situation where there is likely to be a low level of external validity. External validity is further explored in Chapter 6.

Further reading

Hoffmann, T., Bennett, S., & Del Mar, C. (2010). Introduction to evidence-based practice. In T. Hoffmann, S. Bennett, & C. Del Mar (Eds.), Evidence-based practice across the health professions. Sydney: Elsevier.

Straus, S. E., Glasziou, P., Richardson, S. & Haynes, R.B. (2019). Evidence-based medicine: How to practice and teach EBM (5th ed.) Edinburgh: Elsevier Churchill Livingstone.

Liamputtong, P. (2013). Research methods in health: foundations for evidence based practice (2nd ed.). Oxford University Press.

Herbert, R., Jamtvedt, G., Birger Hagen, K. & Mead, J. (2011). Practical Evidence-Based Physiotherapy (2nd ed.). London: Elsevier Health Sciences UK.

Chapter 2

Asking questions to guide your search for evidence

Introduction and learning outcomes

In the following video Simon Pampena gives an overview of searching for evidence:

https://doi.org/10.26181/5c0f4e492ae07

Chapter 1 introduced the five-step approach to EBP, summarised as:

Step 1: Ask an answerable practice-related question

Step 2: Acquire relevant evidence to answer the practice-related question.

Step 3: Appraise the acquired evidence.

Step 4: Apply the appraised evidence to practice.

Step 5: Assess your own performance in executing Steps 1–4, and set learning goals to improve your future performance.

In the following video, Dr Shane Erickson introduces the systematic process of searching for evidence:

https://doi.org/10.26181/5c118fed9fc74

This chapter goes further into the specific process of asking an answerable practice-related question (i.e. Step 1). The five-step EBP process begins with you recognising that you need some information. This information might relate to an intervention, a diagnosis, the **aetiology** of a condition or clients' experiences of illness. An important step is to turn your need for information into an answerable question that you can investigate. It is vital that this first step is completed accurately, because the practice-related question you develop will significantly affect Steps 2–5.

As you will have seen in the introductory video by Dr Shane Erickson for this topic, one method for developing specific, answerable practice-related questions is to use the '**PICO**' mnemonic:

P = population or a clinical problem of interest (or both).

I = intervention you are interested in (this could be an exposure, test, prognostic factor or treatment).

C = comparison (what you think the intervention is better or worse than, if relevant).

O = outcome of interest for your client.

Whether you are a student or a health professional, your time is valuable, and you do not want to waste it undertaking searches that do not yield useful results. Also, you do not want to miss critical information that could lead you to make a poor decision for your patient. You will find that the **PICO** method is an efficient and accurate way to find all of the relevant evidence.

KEY LEARNING OUTCOME

Construct well-structured and answerable practice-related questions (i.e. Step 1 of the five-step approach).

ENABLING OUTCOMES

Once you have completed this chapter, you should be able to:

- identify each of the PICO elements within a practice-related scenario
- develop a practice-related question using a structured approach; for example, the 'PI(C)O' model (the 'C' is in brackets because it is the one element of a practice-related question that is optional)
- evaluate practice-related questions for their appropriateness for addressing a practice-related problem
- identify key features of well-written and poorly written practice-related PICO questions

What is a practice-related question?

As a health professional you will be using your newly learned skills in questioning and searching, to ensure that you can answer questions to find evidence that will help you in managing your patients. Just like the amount of information found through an internet search, the clinical literature is enormous – some estimates suggest that thousands of new studies are published hourly. **Randomised controlled trial**s (RCTs), which are just one type of study and make up only a small proportion of the research found on the average health database, are published at a huge rate. Nevertheless, as a health professional, to help your patients you need to be able to find the vital 'needle' in the overwhelming 'haystack of evidence'.

As explained above, Step 1 is to ask a relevant, answerable practice-related question. The question must be clearly and unambiguously worded. This will allow you to search for the best available evidence to answer that question, and thus to better understand, predict and interpret the results of tests; identify the best way to provide treatment; or find the answers to a patient or client's questions.

As practitioners of evidence-based health care, you are encouraged to ask questions. Being inquisitive means that you are more likely to seek, for example, ways to be more efficient or more effective. Yet being inquisitive does not come naturally to a lot of people; thus, some health professionals find that they need to improve their skills in asking practice-related questions and seeking information to answer those questions.

Types of questions

In your daily practice as a health professional, the questions you ask are likely to fall into one of two types:

> **Background questions** – that is, questions intended to elicit general knowledge about a condition.
>
> **Foreground questions** – that is, questions intended to elicit specific knowledge about managing a patient's specific circumstances.

These two types of question are discussed below.

BACKGROUND QUESTIONS

Background questions target general knowledge that helps you to better understand a condition, assessment or procedure. Examples of background questions are:

> What areas of the brain are involved in complex problem solving?
>
> What causes stroke?
>
> How is sound transmitted through the ear?

You will frequently ask these types of questions as students and new graduates, as you endeavour to learn, for example, about new conditions, assessments and techniques. As you become more experienced and familiar with the areas you work in, you may ask fewer of these background questions, and more questions specifically related to the management of your clients (i.e. foreground questions).

FOREGROUND QUESTIONS

Foreground questions address specific knowledge that will inform clinical decisions and actions. You heard an example of a foreground question in the introductory video for this topic. Another example is the question:

> Is early intensive treatment that targets communication effective for adult stroke patients?

Clearly, this question is looking at the *effectiveness* of early intensive treatment techniques that target communication in adults who have had a stroke. An effectiveness question is one of the five main types of practice-related foreground questions you are likely to ask as health professionals. The five types of question are:

> **Effectiveness** questions.
>
> **Prevention** questions.
>
> **Assessment** questions.
>
> **Description** questions.
>
> **Risk** questions.

Using stroke as the condition of interest, here is an example of each of these five types of practice related questions:

> **Effectiveness** – Is bed rest more effective than exercise in improving the mobility of adults who have had a stroke?
>
> **Prevention** – Does reducing high blood pressure to normal levels prevent strokes in adults?
>
> **Assessment** – Is picture naming an effective method of assessing the language function of adults who have had a stroke?
>
> **Description** – In comparing adult females who are smokers with those who are non-smokers, which group is more likely to have had at least one parent who smoked?
>
> **Risk** – Are 'mini-strokes' in elderly people a risk factor for a more severe stroke in the future?

PICO questions

As discussed above, your practice-related questions need to be well targeted, to ensure that you find the most relevant information to inform your practice. The **PICO** mnemonic will help you to build a question using the most appropriate and important information, which will in turn help you find the most relevant evidence efficiently. You will learn more about searching for evidence in this chapter.

POPULATION AND PROBLEM

The first step in developing a well-built question is to identify the **population** and the **patient's problem** (i.e. the 'P' in PICO). This should include the primary problem, disease or coexisting conditions; for example 'In pre-school aged children who stutter …' sometimes it may also be important to specify the age and gender of a client, if that is relevant to the diagnosis, prognosis or intervention; for example, 'In young adult women with multiple sclerosis…'

When identifying the P in PICO it is helpful to ask the following:

> How could you describe a group of people with a similar problem to your client?
>
> How would you describe the client to another student or colleague?
>
> What are the important characteristics of this patient, and should these characteristics be considered in the search for evidence:
>
> - primary problem
> - patient's main concern or complaint
> - disease or health status
> - age, sex, previous ailments and current medications?

The example of a 'description' practice-related question above was, 'In comparing adult females who are smokers with those who are non-smokers, which group is more likely to have had at least one parent who smoked?' In this example, the *population* is 'adult females'; thus, it includes the age (adult) and the sex (female). However, differentiating whether the adult females are smokers or non-smokers is also critical to the question, and is considered under the intervention and comparison elements, which are discussed below.

INTERVENTION

The second step in developing a question is to identify the **intervention** (i.e. the 'I' in PICO). The term 'intervention' should be considered here in its

broadest sense. It is important to identify *what you plan to do* for that client or the *factor of interest* that you want to find out about (e.g. the use of a specific test, treatment, medication, product or procedure). However, because the intervention may be the main factor you are interested in for that patient, it is important to not just think of an intervention as something that *you* implement as a health professional. For example, in the smoking practice-related question given above, the fact that we want to compare adult females who smoke with adult females who do not smoke means that *smoking* is the factor of interest. Therefore, smoking is considered the 'intervention' – this may seem a little confusing, given that no-one thinks of smoking as an intervention!

The effectiveness question presented above was, 'Is bed rest more effective than exercise in improving the mobility of stroke patients?' In this question, the intervention is 'bed rest', because this is the treatment that we are planning to use. The risk question above was, 'Are 'mini-strokes' in elderly people a risk factor for a more severe stroke in the future?' In this case, the intervention is 'mini-strokes', because this is the factor we are interested in studying; and again, this is a type of practice-related question in which the intervention is not a treatment.

COMPARISON

The third step in developing a question is to identify the **comparison** (i.e. the 'C' in **PICO**); the comparison is the main alternative you are considering, and it should be specific and limited to one alternative choice, to facilitate an effective search. The comparison is the only *optional* component in the PICO question. There will be many occasions when you only want to look at the intervention and do not want to explore alternatives – indeed, in some cases, there may not be an alternative. Sometimes, the comparison you are interested in may be the usual or standard care or treatment; for example, when you want to compare a new treatment technique (the intervention) with the treatment technique you are currently using (the comparison).

The effectiveness question presented above was, 'Is bed rest more effective than exercise in improving the mobility of adult stroke patients?' For this question, the comparison is 'exercise', because that is what we want to compare to the intervention (which we previously identified as being 'bed rest'). However, a question such as the assessment example provided above – 'Is picture naming an effective method of assessing the language function of an adult stroke patient?' – includes an intervention (picture naming) but no comparison. In this case, we are only interested in finding out about picture naming as a method for assessing language function, and not in directly comparing it to anything else.

OUTCOME

The final aspect of the question to be identified is the **outcome** (i.e. the 'O' in **PICO**). The outcome specifies the result or results of what you plan to accomplish, improve or affect, and it should be measurable. Outcomes may consist of:

- relieving or eliminating specific symptoms
- improving or maintaining function
- preventing specific conditions

Being specific with your desired outcomes will yield better search results, and thus will make it easier to find the studies that focus on the outcomes you are searching for. For some outcomes, you may also need to specify whether you are interested in *increasing* the amount of the outcome (e.g. increasing the score on a functional assessment) or *decreasing* it (e.g. reducing pain).

In some cases, determining the outcome of interest may be straightforward; for example, the primary concern of most mothers who bring their stuttering child to a speech pathologist is to reduce the child's stuttering. However, in most areas of evidence-based health practice, shared decision-making is important. Therefore, wherever possible, your patient should be involved in determining the goals of that intervention that are most important to that person, and in many situations the outcome component of your PICO question will be determined by your client's preferences. Identifying the PICO elements in any clinical scenario can be challenging, therefore reviewing examples and practicing make this process easier.

Table 2.1 provides examples of PICO questions broken down into their individual components. You will see that several of the questions do not include a comparison.

Table 2.1 Examples of PICO questions

	Population and problem (P)	What you might do: intervention (I)	Alternate course of action: comparison (C)	What you want to accomplish: outcome (O)
Effectiveness	In people who have had a stroke	is home-based rehabilitation as effective as	hospital-based rehabilitation	in improving ability to perform self-care activities?
Prevention	If high school teachers who are at risk for vocal nodules	undertake seminars on vocal hygiene	or seminars on stress-reduction techniques	will they show a reduced incidence of nodules?
Assessment	In older people living in the community	does the Mini-Mental State Examination		accurately detect the presence of cognitive impairment?

Table 2.1 Examples of PICO questions (continued)

	Population and problem (P)	What you might do: intervention (I)	Alternate course of action: comparison (C)	What you want to accomplish: outcome (O)
Description	What proportion of professional footballers	who have previously suffered a lower back injury		suffer subsequent hamstring injuries?
Risk	In pre-school aged children	is a family history of stuttering		a risk factor for stuttering?

IDENTIFYING THE PICO ELEMENTS IN PRACTICE-RELATED QUESTIONS

Now it is your turn to see whether you can identify each of the **PICO** elements in some examples of practice-related questions. Read the questions and then fill in the table.

Practice-related Question 1

Does reading recovery lead to greater improvements in the reading skills of 6-year-old school children with reading difficulties when compared with no intervention?

Complete the table below based on this question:

Population and problem (P)	What you might do: intervention (I)	Alternate course of action: comparison (C)	What you want to accomplish: outcome (O)

Practice-related Question 2

In children with pain and fever, how does paracetamol compared with ibuprofen affect levels of pain and fever?

Complete the table below based on this question:

Population and problem (P)	What you might do: intervention (I)	Alternate course of action: comparison (C)	What you want to accomplish: outcome (O)

Practice-related Question 3

In elderly people who are unsteady on their feet, is a hip protector a useful preventative measure to reduce the risk of falls?

Complete the table below based on this question:

Population and problem (P)	What you might do: intervention (I)	Alternate course of action: comparison (C)	What you want to accomplish: outcome (O)

Practice-related Question 4

If adults with aphasia are given a language sampling procedure or a verbal fluency rating on the BDAE, which measure correlates best with patient self-perception of communication skills?

Complete the table below based on this question:

Population and problem (P)	What you might do: intervention (I)	Alternate course of action: comparison (C)	What you want to accomplish: outcome (O)

CLINICAL SCENARIOS AND PICO

The critical skill regarding **PICO** questions is being able to turn a clinical scenario into an answerable question that can lead to a search for evidence. Sample scenario 2.1 is an example of how you can turn a clinical scenario involving a friend into a clinical question that would ultimately allow you to search for evidence to help your friend.

Sample scenario 1

You are at a party and a friend tells you that they have discovered a new remedy for a hangover – a tablet called Alcodol. The recommended dosage is two capsules before drinking, and one after, 'if required'. You inform your friend that the easiest way to avoid a hangover is simply not to drink too much, but then you decide that her remedy might be worth further investigation. Given you are learning about the importance of using evidence to inform practice, you want to know if there is any evidence for the effectiveness of Alcodol tablets.

Table 2.2 shows the PICO elements for this scenario.

Table 2.2 PICO elements for determining the effectiveness of Alcodol

Population and problem (P)	What you might do: intervention (I)	Alternate course of action: comparison (C)	What you want to accomplish: outcome (O)
Adults who drink alcohol	Alcodol tablets	hospital-based rehabilitation	Reduced hangover symptoms

The specific wording or ordering of a **PICO** question is flexible, but the question must include all the PICO elements. For this example, a possible question in this case would be, 'Do Alcodol tablets reduce the symptoms of a hangover in adults who drink alcohol?'

WRITING PICO PRACTICE-RELATED QUESTIONS

You need to be able to recognise the differences between 'good' and 'not so good' practice-related questions. Remember that a poorly written question could lead to you finding a lot of irrelevant evidence, or missing important evidence, and thus could make your job of helping your patient more difficult. Once you have written a practice-related question, you should review it to ensure that:

- it includes each of the relevant PICO elements based on the applicable situation
- each of the included PICO elements is written as clearly as possible; for example, often you might need to be specific with the age group you are

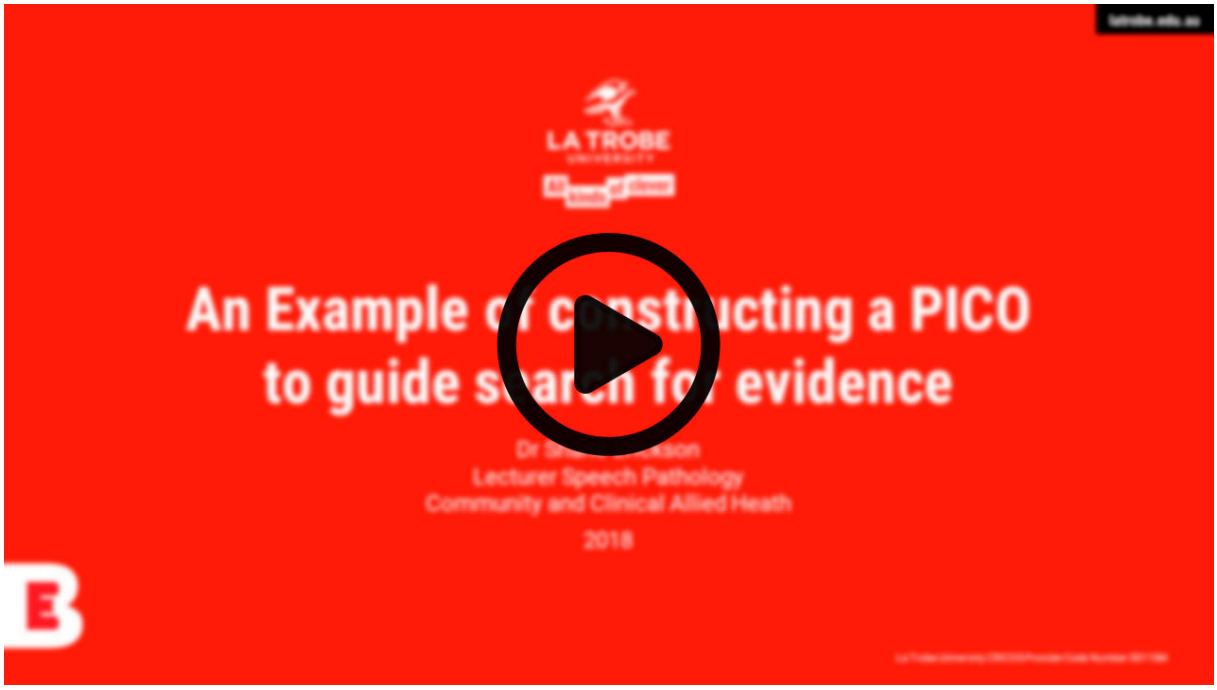
interested in (e.g. pre-school children versus children, or elderly adults versus adults)

it is written as an answerable question (i.e. it finishes with a question mark!)

it is as succinct as possible

it does not include any ambiguous terms

The following video provides an example of how to construct a **PICO** for interventions for back pain:



https://doi.org/10.26181/5c118ff77d2b3

Now it is your turn to transform two sample scenarios into practice-related PICO questions. Whatever the nature of the scenario, you can use the PICO format to create a practice-related question. As explained here, you should then review each question to ensure that it fits the criteria for a well-written PICO question.

Sample scenario 2

After developing a nasty case of athlete's foot, you visit a doctor for an appropriate treatment. You are prescribed a course of Canestan cream (an anti-fungal medication). However, when you tell your mother about your situation, she advises that you do not need to use Canestan, because simply cleaning and drying your feet would work just as well in reducing the symptoms.

You want to find out what the evidence is for these two options for treating the symptoms of athlete's foot. Complete the table below to develop a suitable question for this scenario.

Population and problem (P)	What you might do: intervention (I)	Alternate course of action: comparison (C)	What you want to accomplish: outcome (O)

Sample scenario 3

Your friend Jo tells you that recently her doctor expressed some concerns about her being overweight. As a result, Jo is now keen to lose some weight, and the first thing she wants to try is dieting. She tells you that she recently heard about high protein diets and low carbohydrate diets, but does not know which is likely to be more effective in helping her lose weight.

As with the athlete's foot scenario, you want to find out what the evidence is for these two options for losing weight. Complete the table below to develop a suitable question for this scenario.

Population and problem (P)	What you might do: intervention (I)	Alternate course of action: comparison (C)	What you want to accomplish: outcome (O)

Further reading

Hoffmann, T., Bennett, S., & Del Mar, C. (2010). Introduction to evidence-based practice. In T. Hoffmann, S. Bennett, & C. Del Mar (Eds.), Evidence-based practice across the health professions. Sydney: Elsevier.

The formulation of research questions (2013) p.27. In. S. Polgar & S.A. Thomas, Introduction to research in the health sciences, 6th Ed., Churchill Livingstone Elsevier.

Chapter 3

Acquiring the evidence

Introduction and learning outcomes

Chapter 1 introduced the five-step approach to EBP. This chapter covers Step 2 of the EBP approach: *Acquire relevant evidence to answer the practice-related question.*

You will learn how to conduct a simple search of health databases to find the best evidence to answer your clinical question. Building on Chapter 2, where you created a search strategy based on the **PICO** model, this chapter explains how to construct search strategies using keywords, to ensure that you retrieve accurate and useful results.

KEY LEARNING OUTCOME

Acquire evidence to answer different types of practice-related questions.

ENABLING OUTCOMES

Once you have completed this chapter, you should be able to:

- develop search strategies relevant to various practice-related questions
- implement those search strategies using different databases to collect evidence to address the questions
- implement simple search strategies that use:
  - linking terms, also known as **Boolean operators** (i.e. 'and', 'or' or 'not')
  - **truncation** to search for words with multiple endings (i.e. 'child', 'children' and 'childhood')
- learn how to reference sources in text and create a **reference list** in the required referencing style (APA 6)

Where are the databases, and which one should I use?

The various databases that can be used for health sciences research are listed by subject area in the 'Databases' tab on your university library home page.

Health sciences practitioners need to know how to search the medical literature to find the best evidence for patient care. Library databases will enable you to find a **citation** or reference, which will often lead you to the full text of a journal article. There are two main types of database:

citation databases contain the citation and sometimes the full text (e.g. Medline, CINAHL and PsycINFO)

full text databases contain the citation and full text of the article (e.g. ProQuest)

The most useful databases for you as a health sciences student will be the international databases **Medline** and **CINAHL**. Although it is fine to use Google Scholar initially, it is important to know how to search Medline and CINAHL, because this is where you will find the best available evidence.

The Medline database, is widely recognised as the premier source for bibliographic and abstract coverage of literature in the biomedical and life sciences. Medline encompasses literature from Index Medicus, and the fields of biological and physical sciences; communication disorders; dentistry, nursing and allied health professions; and population and reproductive biology.

CINAHL, is the authoritative resource for nursing and allied health, providing complete coverage of nursing journals and publications. Topics covered by CINAHL are nursing, biomedicine, consumer health and 17 allied health disciplines.

How do I use keywords?

There are two ways to search the Medline and CINAHL databases: by keyword and by medical subject heading (MeSH). In this chapter, we will focus on ***keyword*** searching.

Because CINAHL and Medline are international databases, with contributions from authors across the globe, your searches need to take into account possible variations in spelling, terminology and clinical descriptions, as follows:

Spelling – there are many common variations in spelling; for example, counselling or counseling, ageing or aging, and paediatrics or pediatrics.

Terminology – there are also many common variations in terminology; for example, car driver or automobile driver; and community health nurse (Australia), health visitor (United Kingdom) or neighbourhood nurse (United States).

Medical terminology – as with general terminology, there are many common variations; for example, cerebral vascular accident or cerebrovascular accident.

HOW DO I START A KEYWORD SEARCH?

Imagine that you want to search for evidence for the question, 'Is bed rest or exercise more effective for the treatment of back pain in the elderly?' Table 3.1 shows the **PICO** format for this question, which can then be used to create the search strategy.

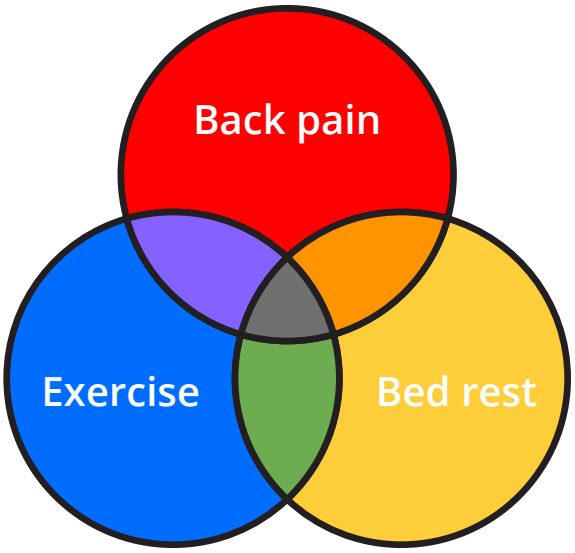
Table 3.1 Using the PICO format to develop keywords for a search strategy

Population (P)	Intervention (I)	Comparative Intervention (C)	Outcome (O)
Back pain / elderly Back Pain OR Backache OR Lower back pain	Exercise OR Physical activity	Bed rest OR Bedrest	Pain relief

BOOLEAN OPERATORS

Based on the PICO map from Table 3.1, we will use **Boolean operators** (i.e. 'and', 'or' or 'not') to search:

- the population group: back pain OR backache OR lower back pain (we will search for the 'elderly' aspect later)
- the concept of exercise OR physical activity
- bed rest OR bedrest.



After searching each concept, we will combine them with the Boolean operator 'and'. Our search results will thus have all three concepts: back pain, exercise and bed rest.

TRUNCATION, WILD CARDS AND PHRASE SEARCHING

In keyword searching, we also use **truncation**, wild cards and phrase searching to ensure that the search is thorough and accurate. Truncation increases the range of search results. In both CINAHL and Medline, the truncation symbol is *, so the term 'child*' will retrieve children, childhood and childless. Hence, in the back pain example given above we could use:

backache* (to retrieve backache or backaches)

low* back pain (to retrieve low back pain or lower back pain)

physical activit* (to retrieve physical activity or physical activities)

Wildcard symbols are used to deal with variations in spelling. In CINAHL, the wildcard symbols used are ? or # (e.g. organi?ation will retrieve both organisation and organization). In Medline, the wildcard symbol used is # (e.g. organi#ation will retrieve both organisation and organization).

Wildcard symbols can also be used where an alternative spelling may contain an extra character. In CINAHL, the wildcard symbol # is used (e.g. p#ediatric will retrieve both paediatric and pediatric, and colo#r will retrieve both colour and color). In Medline, the wildcard symbol **?** is used (e.g. p**?**ediatric will retrieve both paediatric and pediatric, and colo**?**r will retrieve both colour and color).

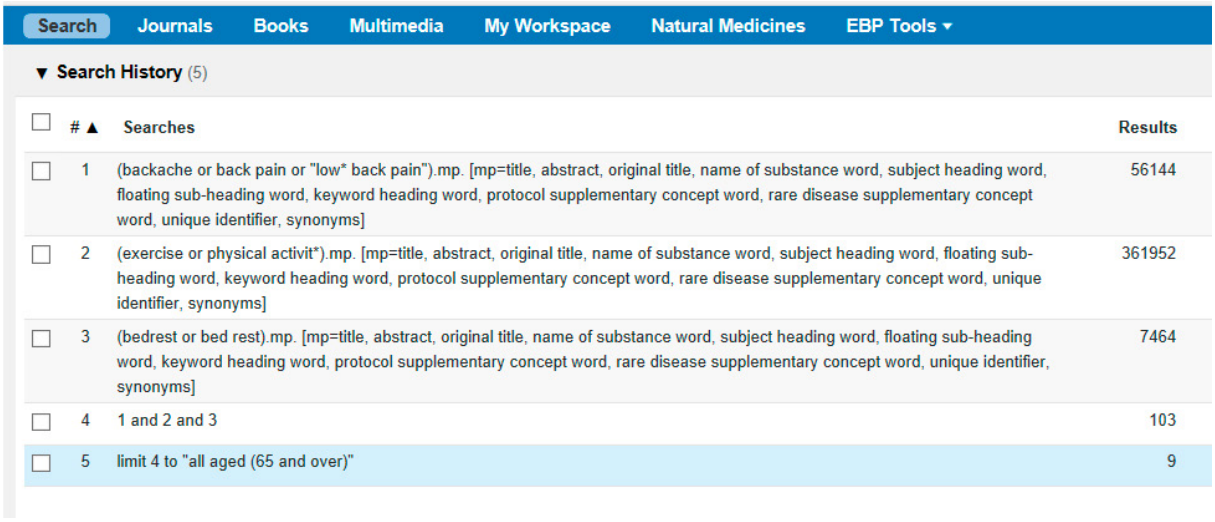
Phrase searching involves using quotation marks to search for words as a phrase; for example, “quality of life” or “acquired brain injury”.

Limiting results

Let us return to our question: ‘Is bed rest or exercise more effective for treatment of back pain in the elderly?’ Once you have completed your search, you can limit the search to ‘elderly’ as a population group in the databases. Underneath the search results screen on the left, there are options to refine your results to **peer-reviewed**, publication date and more. Click on ***Show More*** and scroll down to ***Age Groups*** and select Aged 65+, then search.

You will find that limiting the search in this way reduces the number of results from 103 to 9 in Medline (15.9.2018), as shown in Figure 3.2.

Figure 3.2 Medline search history for the question 'Is bed rest or exercise more effective for treatment of back pain in the elderly?'



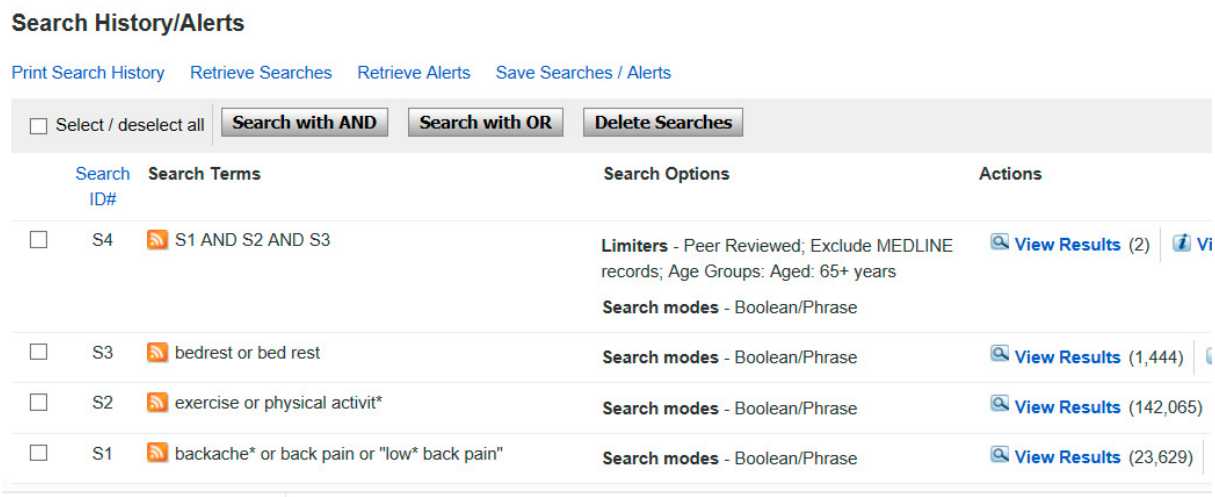
Search | Journals | Books | Multimedia | My Workspace | Natural Medicines | EBP Tools

▼ Search History (5)

# ▲	Searches	Results
1	(backache or back pain or "low* back pain").mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	56144
2	(exercise or physical activit*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	361952
3	(bedrest or bed rest).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	7464
4	1 and 2 and 3	103
5	limit 4 to "all aged (65 and over)"	9

'Screenshot image of the Ovid search history' used with permission under terms of Wolters Kluwer licence

Figure 3.3 CINAHL search history for the question 'Is bed rest or exercise more effective for treatment of back pain in the elderly?'



Search History/Alerts

Print Search History | Retrieve Searches | Retrieve Alerts | Save Searches / Alerts

Select / deselect all | Search with AND | Search with OR | Delete Searches

Search ID#	Search Terms	Search Options	Actions
S4	S1 AND S2 AND S3	**Limiters** - Peer Reviewed; Exclude MEDLINE records; Age Groups: Aged: 65+ years **Search modes** - Boolean/Phrase	View Results (2) | Vi
S3	bedrest or bed rest	**Search modes** - Boolean/Phrase	View Results (1,444)
S2	exercise or physical activit*	**Search modes** - Boolean/Phrase	View Results (142,065)
S1	backache* or back pain or "low* back pain"	**Search modes** - Boolean/Phrase	View Results (23,629)

'CINAHL' screen capture used with permission from EBSCO Information Services

Figure 3.4 Examples of search results for the question 'Is bed rest or exercise more effective for treatment of back pain in the elderly?'

'EBSCO' screen capture used with permission from EBSCO Information Services

The following video presents an overview of keyword searching in CINAHL and Medline:

https://doi.org/10.26181/5c118f7e163eb

https://doi.org/10.26181/5c118f6dbf01f

More help with searching is available from the Health Databases guide.

Finding journal articles

Once you have identified the journal articles that are relevant to your question, you need to find the actual articles. Sometimes, the full text of the journal article is attached to the record in the search results.

If no full text is available, search for the journal article under the *Library Search* box. If you cannot find it there, you can check the journal title under the *Journals* tab on the Library home page; then link through to the year, volume and issue to find the full text of the article.

If the article is not available, Document Delivery Services (under Quick Links on the Library home page) will order you a copy (in 1–3 days), at no charge.

Referencing your sources

You must cite your references in your assessments, to support your arguments with evidence and to protect yourself against charges of

plagiarism. Referencing your sources also shows the breadth of your research into the topic, and enables the reader to locate the sources referred to.

You need to cite your source when you:

- paraphrase someone else's work
- summarise another person's ideas
- use direct quotes
- refer to the ideas or theories of another person's work

The College of SHE at La Trobe University and most health sciences researchers use the APA 6 edition as the referencing style. When referring to another work in an assessment, you need to cite that work in the text and list it in the references at the end of the document. Here are examples of in-text citations for a journal article:

Direct quote – Parents smoking in their cars and in the family home 'should be considered as intermediary factors in the pathway between parental and student smoking' (Glover et al., 2011, p. 1028).

Paraphrasing – Glover et al. (2011) report the factors in the pathway between parental and student smoking.

The **reference list** would then give the full reference as:

Glover, M., Scragg, R., Min, S., Kira, A., Nosa, V., McCool, J., & Bullen, C. (2011). Driving kids to smoke? Children's reported exposure to smoke in cars and early smoking initiation. *Addictive Behaviors, 36*(11), 1027–1031. doi:10.1016/j.addbeh.2011.06.003

Further reading

Greenhalgh, T. (2014). How to read a paper: The basics of evidence-based medicine (Fifth ed.) Chichester, West Sussex John Wiley & Sons.

Wilczynski, N. & McGibbon, A. (2013). Finding the evidence. In T. Hoffman, s. Bennett, & C. Del Mar (Eds.), Evidence-based practice across the health professions, Australia: Elsevier

Chapter 4

The importance of critically appraising research

Introduction and learning outcomes

In the following video Simon Pampena discusses the importance of critically appraising research:

https://doi.org/10.26181/5c0f515f4a2f5

There are a number of contexts in which the ability to critically appraise research is likely to be important. For example, students are often asked to critique a journal article, so that they can develop skills identifying the strengths and weaknesses of published research. Students undertaking research projects may need to critically evaluate research, to decide whether it is suitable to include in a review of the literature related to their topic, or to consider the results of their own research in light of what is already known about the topic. Health professionals need to critically appraise research so that they can make a well-informed decision about how best to care for their client or patient. This decision needs to take into account both the **validity** of the research and whether the findings are clinically important.

The appraisal of research is Step 3 of the five-step EBP process, and hence usually occurs after a student or clinician has asked a clinical question (Step 1) and acquired relevant evidence (Step 2). When you first start to critically appraise research, it can seem quite a daunting task (and sometimes it is daunting even for experienced readers of research!). However, it is clearly important to learn this critical skill. Happily, this is a skill that improves with practice, and having a systematic, logical method to follow makes the task much less daunting! This chapter introduces the basics of critical appraisal; later chapters investigate in more depth the aspects of research that need to be appraised to determine its quality.

KEY LEARNING OUTCOME

Describe the importance of critical appraisal for EBP, and outline the key steps in the appraisal process.

ENABLING OUTCOMES

Once you have completed this chapter, you should be able to:

- describe critical appraisal in the context of a journal article
- provide at least one rationale for why you will need to undertake a critical appraisal of research
- describe how critical appraisal tools can be used to assist in appraising research studies of various designs.

The benefits of critical appraisal

As a health professional, you are expected to make decisions and provide information for your patients or clients using the best available evidence that is likely to have a clinically significant impact. Obviously, this means that you will need some way to determine the quality of evidence, and thus whether it is likely to help your client. The 'quality' component of a research study relates to the way in which the research was conducted. We need to be able to trust that when researchers report findings from their research, those findings are a true and accurate representation and are not due to a biased methodology.

One of the first things to check when reading a research study is whether it has been *peer reviewed.* Typically, before an article is published in an academic journal, it is reviewed by suitably qualified and experienced researchers familiar with the topic that the research relates to (i.e. by peers). Many journals will require several peer reviews before accepting an article for publication. You can check the peer review process for any given academic journal by visiting the journal's homepage on the internet and reading the author submission guidelines. However, peer review in itself does not automatically mean that the research is of good quality. The level of scrutiny undertaken by peer reviewers varies widely among journals.

Two key terms related to critical appraisal are:

Validity – in relation to research conducted using a quantitative methodology.

Rigour – in relation to research conducted using a qualitative methodology.

Chapters 6 and 7 discussed **validity** in relation to quantitative research; chapter 13 discusses what constitutes rigour in **qualitative research**.

Another point to consider is the *applicability* of the research. The way the research was conducted or the nature of the participants in the study might mean that it is hard to generalise the findings of the study to other settings. Chapter 6 explained that in quantitative research this is referred to as 'external validity' (whereas the control of **bias** is referred to as 'internal validity'). Sometimes researchers try so hard to control every possible potential bias or confounding variable to ensure internal validity that they create an unrealistic environment and thus jeopardise the ability to generalise the findings (i.e. the external validity). When you are critically appraising research, keep in mind that ideally we are looking for studies that are both good quality and able to be generalised.

Undertaking a critical appraisal

There are many acceptable ways to critically appraise research; however, simply reading the abstract and discussion is not one of them! Critical appraisal requires a critique of each part of the research that is reported. You might find it helpful to start by briefly reading the entire article at a superficial level to gain a general understanding of the study before reading it several times in detail and making notes. For those new to the process of critical appraisal, you might consider skipping the abstract; once you have read the article a number of times and taken notes, you can then compare your understanding of the research with what is reported in the abstract.

Given that critical appraisal requires a detailed evaluation of all parts of the journal article, and there are many ways in which the quality of the research can be jeopardised, the process can seem a little overwhelming. Thus, it can be helpful to use a published critical appraisal tool, such as one of the following:

> The Critical Appraisal Skills Program (CASP) provides a set of eight critical appraisal tools available as electronic files for systematic reviews, RCTs, **cohort** studies, case–control studies, economic evaluations, diagnostic studies, qualitative studies and clinical prediction rule. Each appraisal tool includes 10 questions categorised under the headings: are the results of the study valid; what are the results; will the results help locally?
>
> La Trobe University critical appraisal guide

The Joanna Briggs Institute provides a set of 13 critical appraisal tools covering a wide range of study designs, including the most common quantitative designs, systematic reviews and **qualitative research**

PRISMA is an evidence based minimum set of items for reporting in systematic reviews and meta-analyses

CONSORT stands for Consolidated Standards of Reporting Trials. It provides a 25-item checklist that can be used to consider how a research trial was designed, analysed and interpreted. Although the CONSORT 2010 checklist is broadly designed for the reporting of randomised trials, there are a further 10 checklists that can be used for specific types of randomised trials, including 'non-inferiority trials' and 'N-of-1' trials

'Non-inferiority trials' are designed to show that a treatment is equivalent to another treatment. Such trials may be relevant when comparing a new treatment with the standard treatment, when the new treatment may offer important advantages in terms of improved safety, convenience, compliance or cost. For example, a trial might compare telehealth delivery of a treatment (i.e. treatment delivered via webcam) with the standard in-clinic delivery of the same treatment. Telehealth treatment typically offers advantages such as reduced costs and increased convenience; however, to make these benefits relevant, the researchers would need to show that telehealth delivery is still equivalent to in-clinic delivery in terms of efficacy.

An 'N-of-1' trial, also referred to as a single **case study**, involves a single participant. Random allocation (as in RCTs) can be used to determine the order in which an **experimental** and a control intervention are given to the participant. Such designs have been used for people with rare health conditions. However, researchers are increasingly recognising the important role that well-designed N-of-1 trials can play in the study of more common health conditions.

Essentially, these critical appraisal tools guide you through a series of questions that evaluate the research in terms of the following questions:

What was the clinical question to be answered by the research?

Which study design was used and was it appropriate?

What were the characteristics of the sample, and what was the recruitment procedure?

What data were collected and how did the data collection occur?

What was the **independent variable** (in quantitative research), and how was it administered?

What other potential sources of **bias** may have affected the study?

What are the results and are they applicable to practice?

You may find that as you develop skills in appraising an article, you will not need to use a critical appraisal tool. This is particularly the case when you read a lot of research from a specific area of practice, and you recognise common quality pitfalls that researchers may fail to address. Until that time, using a critical appraisal tool will help to ensure that you complete a thorough job of what is a really important aspect of the EBP process.

Further reading

Greenhalgh, T. (2014). How to read a paper: The basics of evidence-based medicine (Fifth ed.).

Greenhalgh, T. (1997). How to read a paper: Assessing the methodological quality of published papers. BMJ, 315(7103), 305-308.

Critical evaluation of published research (2013) p.191. In. S. Polgar & S.A. Thomas, Introduction to research in the health sciences, 6th Ed., Churchill Livingstone Elsevier.

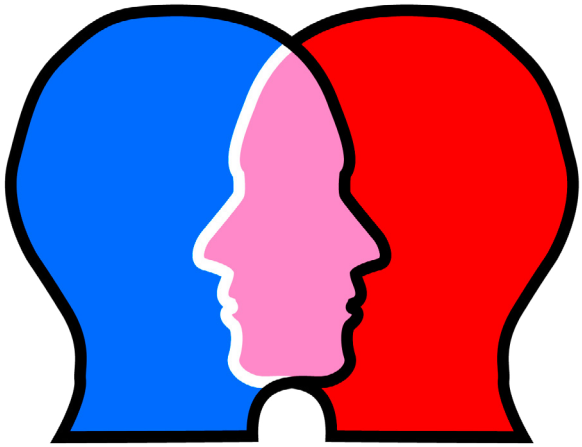
Buccheri, R., & Sharifi, C. (2017). Critical appraisal tools and reporting guidelines for evidence-based practice. Worldviews on evidence-based nursing, 14(6), 463-472

Chapter 5

Introduction to quantitative research design

Introduction and learning outcomes

Chapter 1 introduced the broad categories of *quantitative* and **qualitative research**. It explained that quantitative research seeks to test theories by analysing relationships, and that this type of research involves measuring specific characteristics of the participants in the studies. Chapter 1 also talked about the way an initial observation (in that case about a cat's TV watching habits) resulted in the generation of a testable theory (about whether cats watch more TV when there are birds on the screen). You can make initial observations about many things that happen around you that could lead to theories, which could ultimately be tested scientifically.



Lecturers of students in health degrees often comment that their students are highly empathic, seeming to really understand other people's experiences, feelings and their point of view. Of course, this is just a casual observation and we would need to collect some data to see whether this observation is actually true. To do this, we would first need to work out what the variable is that needs to be measured. Perhaps we could learn about the levels of empathy of students in health courses by measuring their personality, using one of the many well-established questionnaires for this purpose. Let us imagine that we did this, and found that 90% of a large, randomly selected sample of students in health science courses were classified as highly empathic – these data would tend to support the initial observation!

The next step would be to explain these data. One explanation could be that people who are highly empathic are more likely to want to pursue careers that focus on helping other people. This is called a *theory* (in the same way that the idea discussed earlier, that cats prefer to watch TV when it features birds, is a theory). The initial observation about the students was verified by collecting data, and additional data could be collected to test this theory. We could also make predictions from this theory. For example, we could predict that among potential students attending health degree information sessions at La Trobe University Open Days, a greater proportion would be classified as having higher empathy than in the general population. Such a prediction from a theory is known as a **hypothesis**.

The next step is to test the **hypothesis**. We could do this by asking a team of psychologists to assess each potential student at health degree information sessions at La Trobe University Open Days. The psychologists could use this data to rate each potential student's level of empathy, and we could then compare the proportion who are rated as having high empathy with previously established data for the proportion of the general population who have high empathy. The findings from this analysis would either support or refute our hypothesis.

Translating research questions into testable hypotheses is a critical aspect of quantitative research. As health professionals, you will need to identify research that seeks to ask research questions and to test hypotheses relevant to the patients and clients that you work with. This research can take many different forms, depending on the research question being investigated. This chapter looks at features of different types of quantitative research designs, describes the strengths and limitations of these designs, discusses how certain designs best suit certain research questions, and finally looks at the structure of a quantitative journal article.

KEY LEARNING OUTCOME

Outline the defining features, and the strengths and limitations of different types of quantitative research.

ENABLING OUTCOMES

Once you have completed this chapter, you should be able to:

- describe the purpose of hypotheses in formulating research aims
- explain defining features of different types of quantitative research; for example, RCTs, quasi-experimental studies and cross-sectional studies
- identify strengths and limitations of various quantitative designs
- determine which research questions are appropriate for specific quantitative designs
- describe the organisation and structure of a quantitative journal article

Relationships between research questions and research designs

THE RESEARCH PROCESS

Some types of research are designed with observational goals in mind; that is, the researcher gathers data, makes observations and measures **phenomena** (as in the example given above regarding the proportion of health degree students who have high empathy). Other types of research seek to go beyond description, aiming to identify the causes of illnesses, disorders and disabilities. Another major goal of some research is to demonstrate that the interventions used to treat specific conditions actually cause beneficial changes. Finally, some research is intended to test the accuracy of the tools used to diagnose particular illnesses or conditions.

In addition to the many different aims of research, there are many different ways that research studies are carried out to investigate these aims. As you will learn in this chapter, the way a research study is conducted can greatly affect whether we can have confidence in the results.

Given that different types of research questions are best addressed by different types of research studies, we will consider the best type of study to answer a specific question. In Chapter 2 (when you learned to formulate practice-related research questions using the **PICO** approach), you were given the scenario in which a friend was preaching the benefits of Alcodol tablets to reduce the symptoms of a hangover. Imagine that you have completed a search of the literature, and found no scientific evidence to support the use of Alcodol tablets. You tell your friend that this is the case and she (quite correctly) states that this does not mean that the tablets do not work, it just means that they have not yet been scientifically evaluated. With this in mind, you start thinking about the best way to scientifically test the efficacy of Alcodol tablets in reducing the symptoms of a hangover.

One of the first steps in developing this research study would be to construct a **hypothesis**. Most hypotheses can be expressed in terms of two **variables**: a proposed *cause* and a proposed *outcome*. If we use the scientific statement, 'Alcodol tablets effectively reduce the symptoms of a hangover', then the proposed cause is 'Alcodol tablets' and the proposed effect is 'reduced hangover symptoms'. Variables can be *independent* or *dependent*:

> a variable that we consider as being a cause is an **independent variable**, because its value does not depend on any other variables – in this experiment, the Alcodol tablets are the independent variable
>
> a variable that we consider as being an effect is a **dependent variable**, because its value depends on the cause (i.e. the independent variable) – in this experiment, the hangover symptoms are the dependent variable.

In experiments seeking to establish the relationship between cause and effect, the researcher manipulates the **independent variable**, then measures the effect on the **dependent variable**. Thus, in this example, you would need to manipulate the Alcodol tablets and measure the effect on the hangover symptoms. You will learn about the important aspects of measurement in Chapter 7.

The best way to set up an experiment to test the efficacy of Alcodol tablets would be to design an RCT. The goal of an RCT is to try to ensure that any observed effects are the result of the intervention and not some other factor. A sample of study participants is drawn from a population, and each participant is assigned, by a random method, to either an intervention group or to a **control group**. Researchers aim to create two groups that are as close as possible in terms of the participants' characteristics (e.g. age, sex and any particular characteristics that are important to the trial).

To study the effectiveness of Alcodol, there are many additional factors that you would need to consider before starting the trial. Such factors include the amount of alcohol each participant should consume, the amount of sleep they have, and whether they have something to eat after drinking. Ideally, once you have established these factors and ensured that they are evenly distributed across the two groups, you would give the intervention group the intervention (i.e. the Alcodol tablet). The control group would need to receive a **placebo** tablet that looks exactly like an Alcodol tablet but is actually a sugar tablet with no expected effect – this would address any expectation the participants had about the intervention working. Ideally, both groups would also be blind to (i.e. unaware of) whether they received the Alcodol tablet or the placebo. We look more closely at the use of placebos and **blinding** later in this chapter. The measurement of the **dependent variable** (i.e. hangover symptoms) would occur at a designated point (e.g. at 9 am the next day), and the final step would be to analyse and interpret the results.

Placebo effect

The placebo effect does not tend to result in a cure of a health condition; instead, it tends to relieve the symptoms (e.g. perception of pain). Our understanding of the mechanism of this effect is that the expectation that a treatment will lead to an improvement may result in changes to the neurochemistry of the brain that lessen the severity of symptoms. Hence, researchers wanting to show cause and effect must rely on more than just the experiences of an individual who has undergone the treatment of interest, because they cannot determine whether any changes are due to the intervention itself or to the placebo effect. The person may have experienced the same relief in symptoms to the same extent and in the same timeframe, simply because of the expectation of the positive effects of the treatment.

MATCHING STUDY DESIGNS TO RESEARCH QUESTIONS

Randomised controlled trials (RCTs)

RCTs are useful in testing the efficacy of interventions, and (at least in theory) in determining the causes of some diseases and health conditions. In practice, however, RCTs are limited in their usefulness for investigating the causes of diseases and health conditions. This is partly because it is usually unethical to expose the intervention group to the factor or factors that are suspected to be the cause of the disease or health condition. For example, to investigate whether cigarette smoking causes lung cancer, it would be unethical to randomly allocate participants to an intervention (smoking group) and a control (non-smoking group), and then have the smoking group smoke a packet of cigarettes each day for an extensive designated period of time.

Prospective cohort studies

Given the limitations of RCTs, the evidence that smoking causes lung cancer has largely come from different types of research studies; in particular, from prospective **cohort** studies. Such studies on lung cancer involved following a large group of people (including smokers and non-smokers) over time, and determining the number of cases of lung cancer in the group. This is the essence of a *prospective cohort study*; a group of participants (the cohort) is identified and then followed over time, to determine who is exposed to a potential causal factor for a disease or health condition of interest, and who develops the disease or health condition. The 'prospective' part refers to the fact that the participants are followed prospectively (i.e. into the future).

In the same way that RCTs are the best way to answer research questions relating to interventions, prospective cohort studies are the best approach for questions relating to the cause of a disease or health condition (i.e. **aetiology**). Also, because prospective cohort studies follow people with a specific disease or health condition over time, and measure outcomes as they happen, this study design is best at answering research questions related to the prognosis (i.e. the likely outcome) of a disease or health condition.

Researching diagnostic tests

As mentioned above, another focus of research questions is determining the accuracy of the assessments used by health professionals to diagnose diseases or health conditions. The best design for this type of research involves testing a large sample of people with both the new diagnostic test and an established test, and comparing the results. For example, the police have recently introduced new breathalysers that only require drivers to speak into them (usually by counting to five) to measure their blood alcohol content (BAC). To establish the accuracy of this type of assessment, researchers are likely to have compared the BAC readings on these new machines with the BAC readings taken on the 'tried and true' breathalysers that drivers had to blow into using a straw. By comparing these two assessments independently, researchers could ascertain the diagnostic accuracy of the new technology.

Summary

Table 5.1 summarises the type of research questions that might be asked, and the best type of study or research design in each case.

Table 5.1 Matching research question to type of study or research design

Type of question	Best type of study or research design
Therapy or treatment	Randomised controlled trial
Diagnosis	Prospective, blind comparison to a gold standard
Aetiology	Prospective cohort study
Prognosis	Prospective cohort study
Prevention	Randomised controlled trial

In the following video Dr Elly Djouma provides an example of a quantitative research approach in the context of substance abuse in rats:

https://doi.org/10.26181/5c119021a2c88

Other factors affecting choice of study design

In general, there are two key considerations for researchers when selecting the research design for a particular study:

> The **first** consideration is the type of research question that is being asked; that is, is it an intervention question or an **aetiology** question?
>
> The **second** consideration is the need to maximise internal validity. Chapter 6 introduces the notion of internal validity; that is, the degree of certainty that we can have about the correctness of conclusions drawn from the study's findings. We will look closer at internal validity below.

The preferred study design for some common types of research questions was summarised in Table 5.1. However, in practice, it is not always possible to use a particular design, perhaps because the conditions required for a certain type of research are not feasible or appropriate. For example, we know that RCTs are the most appropriate way to test the effectiveness of a treatment. In the proposed Alcodol study described above, the nature of the trial meant that we could construct an RCT in which the participants could be divided into two groups: those who received the Alcodol tablet and those who received a **placebo**. However, if you consider a study in which the participants are critically ill and the researchers want to test a medication that is likely to save their lives, it would not be ethical for half of those participants to receive a placebo (although it might be ethical for the **control group** to receive the standard treatment while the test group received the **experimental** treatment – if the participants were randomly allocated to the groups, this could still be considered an RCT). Researchers seek to use the best research design to answer their particular question that is also feasible once pragmatic, ethical and economic considerations have been taken into account.

Rigorous study designs such as RCTs (the most rigorous design) can be expensive, time consuming and difficult to coordinate. Therefore, when little is known about an issue such as a new treatment, a more exploratory method is often the appropriate first step. Findings from exploratory research can provide the rationale for potentially large amounts of time and money being invested into a research project. As the researchers' level of knowledge about an issue increases, study designs become more rigorous, particularly once most **variables** that could influence the outcome are understood and can be controlled by the researcher.

In cases where particular research designs are not feasible (perhaps because of ethical or logistical reasons), researchers can use different designs to address research questions. However, the choice of design will affect the

level of confidence we can have in the results of studies, because some designs have inherent limitations. For example, if there is no **control group** to compare with the treatment group, we cannot be sure that the outcomes of a study are due to the treatment alone. Other factors that could affect outcomes include disease progression, medication use, lifestyle and environmental changes.

Other common research designs

There are many other possible research designs; this section describes some of the most common.

Quasi-experimental design

Occasionally, randomisation is not possible because of ethical or practical reasons; in these circumstances, an alternative is quasi-experimental research. Such research is identical to an RCT in most respects, but the participants are not **randomised** to treatment groups. This lack of random allocation can introduce **bias** related to the way in which participants are allocated to each group. Additionally, if the groups are not matched closely enough on key characteristics, it can be difficult to make accurate comparisons between the treatment and **control group**s.

Case-controlled studies

Case-controlled studies also involve an intervention group and a comparison group. However, they differ from the quasi-experimental design in that they are often retrospective (i.e. looking at an issue after it has happened). They rely on the identification of a group of people with an outcome or disorder of interest being compared retrospectively with a control group who do not have that outcome or disorder. A case-controlled study is a relatively inexpensive way to explore an issue. However, there are many potential problems – mainly associated with the accuracy of retrospective data – that make it difficult to conclude what factor or factors are responsible for the outcomes of such a study.

Before–after studies

A before–after design is generally used to evaluate a single group of clients who receive a treatment. Information about the initial status of a group of clients, in terms of the outcomes of interest, is measured before treatment is received and again after treatment. This type of design is useful when researchers do not want to withhold treatment from any clients. However, because there is no **control group**, it is impossible to judge whether the treatment alone was responsible for any changes in the outcomes.

Cross-sectional studies

A cross-sectional design involves a single group of people, with the evaluation of the whole group carried out at one point in time. **Survey**s, questionnaires and interviews are common **methods** used in cross-sectional studies. It is difficult to draw cause–effect conclusions from the results of such studies, because it is impossible to know whether all factors have been included in the evaluation. Additionally, such studies often ask participants to recall events that have happened in the past**, which** may decrease the accuracy of this information.

Single-case design

A single-case design involves one client, or a number of clients, being followed (as individuals, not as a group) over time. The key feature is the evaluation of clients for the outcome or outcomes of interest, both before (i.e. at baseline) and after the intervention. This design allows an individual to serve as their own 'control'. However, it is difficult to conclude from a single-**case study** that the treatment alone caused any changes, because other factors (e.g. disease severity) may change over time. Also, participant numbers are low, making it hard to generalise beyond the person or people in the study. A single-case design can, however, be useful when the population of people with a particular diagnosis is small.

Methodological quality

The strength of research evidence depends not just on the study design used, but also on the methodological quality of the study (i.e. how the study was conducted). For example, the evidence about an intervention gained from an RCT that is poorly designed and conducted is likely to be weaker than evidence about the same intervention gained from a **cohort** study that is well-designed and conducted. Chapter 6 provides more information about **bias** and the ways in which it can affect research studies.

As practitioners of evidence-based medicine, you will need to become skilled at spotting **bias** in research, and at determining how much influence it has had on the strength of the evidence you are reviewing. In our imaginary Alcodol study, you might remember that we gave the **control group** a **placebo** tablet instead of the Alcodol tablet. We did this to account for the fact that participants might have an expectation that taking Alcodol will reduce hangover symptoms; that is, we gave a placebo to reduce the likelihood of participants reporting more positive effects than actually occurred. A strategy such as a placebo helps to increase our confidence that the reported findings from a research study are true.

Basic organisation and structure of a quantitative research study

Before you start reading and appraising quantitative research, you need to be familiar with the structure and organisation of a journal article. Fortunately, the structure is largely consistent across journals that publish academic research. Thus, the more you read, the easier it is to interpret and appraise an article. Once you are aware of the structure of a journal article, you will be able to quickly find important details about the study, including any potential flaws in the research.

A typical journal article dealing with quantitative research has the following structure (See **Annex 2**):

Title – Informative, attract the reader's attention, should accurately reflect the nature and focus of the study.

Abstract – Short summary, provides an overview of what the research is about, what was done, how it was done, what was found, and what the results mean

Keywords – 6-8 keywords used to draw the reader's attention, also used to locate articles in electronic databases.

Introduction – Brief overview of previous relevant research, provides a rationale for the study and an outline for what the research is aiming to do. Authors highlight a gap in knowledge and describe what their study will provide in relation to this gap.

Methods – Summarises the procedure, providing enough detail that another research study could replicate it including: participants, materials, study design, procedure and the process of data collection and analysis.

Results – Summarises the data collected and statistical analyses performed. Should report the results without any type of subjective interpretation. Some research intends only to describe the results for the sample, while other research attempts to make inferences about the population from the sample.

Discussion – Summarises and interprets findings, relates the findings back to previous research, considers the original research question or hypothesis, and discusses the clinical implications for the client and the profession.

Conclusion – Provides any limitations of the research and recommendations for future research.

Further reading

Liamputtong, P., & Bondas, T. (2016). Research Methods in Health (3rd ed.). Melbourne: OUPANZ.

Polgar, S. & Thomas, S.A. (2013). Introduction to research in the health sciences, 6th Ed., Churchill Livingstone Elsevier.

Chapter 6

Internal and external validity

Introduction and learning outcomes

This chapter discusses the term '**validity**' in the context of the integrity of the conclusions that are generated from research; that is, whether a causal relationship can be inferred (internal validity), and whether the results of a study can be generalised to a broader context (external validity). This chapter focuses on internal and external validity; chapter 7 looks at measurement validity.

In quantitative research, the researcher needs to be aware of **bias** and how this may affect the credibility of findings. Researchers investigating the effect of a treatment are usually trying to establish that observed effects can be directly attributed to the treatment itself, and not to some other extraneous variable. They will also take steps to try to ensure that their findings can be generalised beyond just the participants in the study. As consumers of research, we need to be able to identify the steps taken by researchers to maximise internal and external validity, and then decide whether we trust the findings of the research sufficiently to inform decisions in our practice.

KEY LEARNING OUTCOME

Identify biases from research evidence and describe how they threaten internal and external validity.

ENABLING OUTCOMES

Once you have completed this chapter, you should be able to:

- describe the concepts of internal and external validity
- identify specific types of **bias** associated with internal and external validity
- explain how common types of bias can be avoided

Validity

Validity concerns the integrity of the findings that result from a particular study. As a health-care practitioner, when you are reviewing journal articles you will need to be aware of the main types of validity so that you can assess how valid the reported findings are.

INTERNAL VALIDITY

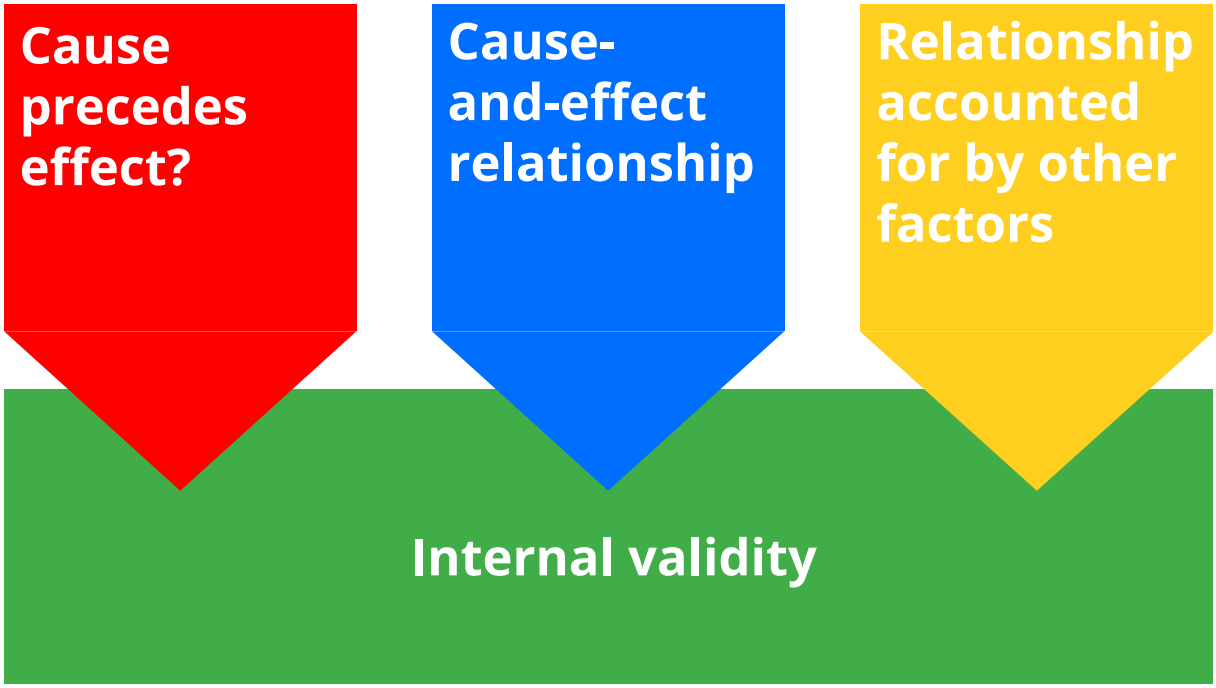
Internal validity is concerned with the question of causality; that is, whether we can be confident that we can draw a causal relationship between two **variables**. In discussing issues of causality we need to understand the difference between a dependent and an **independent variable**:

> The **dependent variable** (symbolised by y) is the variable that researchers either need to understand, or be able to explain why it varies.
>
> An **independent variable** (symbolised by x) is a variable that researchers believe may produce some variation in the dependent variable. If we suggest that x is responsible for the variation in y, we need to be confident that this is the case, and be sure that the variation in y is not due to some other extraneous variable.

To give an example, suppose we ran birth control classes that focused on teaching adolescents in a particular population about different birth control options. If we then found that the adolescents in that population had increased knowledge of birth control, could we have confidence that this finding was a direct result of the adolescents' participation in the birth control classes, or were other factors responsible? To be confident that the independent variable (the birth control classes) was responsible for the dependent variable (knowledge of birth control options), we would need to attempt to eliminate other plausible explanations for the finding.

Figure 6.1 demonstrates the factors that need to be considered in establishing internal validity. While the goal of a research study, like the hypothetical birth control trial, may be to show a cause and effect relationship, researchers interpretation of results can be affected by the logical fallacy that because Y (knowledge of birth control options) followed X (the birth control classes), Y must have been caused by X. In Latin this is referred to as post hoc ergo propter hoc or post hoc for short. Other confounding factors (i.e. perhaps the participants sought further information in addition to attending the classes) might actually be responsible for changes detected and this affects the internal validity of the study.

Figure 6.1 Internal validity



EXTERNAL VALIDITY

Internal validity can be contrasted with external validity, which relates to the degree in which the study's findings can be generalised beyond the people involved in the study. After all, most research is intended to guide practice, and if the results cannot be generalised beyond the small group of people in the study, then the real value of the findings is limited. Essentially, external validity asks whether causal relationships can be generalised to different measures, persons, settings and times.

A common criticism of treatment studies is that, by virtue of trying to increase internal validity (i.e. by designing a study that is highly controlled), the researchers only show that a treatment is effective in ideal circumstances. Therefore, when clinicians read such studies they sometimes question the usefulness of the findings because the setting does not reflect their workplace. This is an example of ecological validity i.e. do the findings of a research study reflect real life settings?

Historically, however, researchers have tended to focus on maximising internal validity (often to the detriment of external validity) because of the perception that it is more important to show that a treatment is effective in ideal conditions than to show that it is effective with different populations and in different settings. This focus has probably also been influenced by academic journals and funding organisations that require researchers to demonstrate high internal validity if the research is to receive funding or be accepted for publication.

Often research is divided into levels:

Efficacy studies – these focus on demonstrating internal validity through a highly controlled methodology.

Effectiveness studies – these follow efficacy studies and are carried out in less controlled situations that are closer to real life.

To establish external validity, we need to consider how well the research, using a sample, can be generalised to the population as a whole. The following are some key aspects of the sample that should be evaluated:

How many participants were involved, and was this a sufficient number to generalise to the population?

Was the sample size justified? Preliminary studies may involve small sample sizes but if the researchers aim to demonstrate efficacy, then a larger sample is needed.

Is a clear description of the key characteristics of the sample provided? For example, the age and gender, and the time of onset of the health condition of interest.

If there were multiple groups in the study, were they comparable in terms of size and participant characteristics?

Were appropriate inclusion and exclusion criteria described?

Identifying common types of bias and their impact on research

This section introduces some common types of **bias** and how they may affect the internal and external validity of research. As previously mentioned, internal validity is related to the extent that we can identify a causal relationship between the independent and **dependent variable**. This is particularly important in health research when we need to identify causal relationships between interventions and outcomes. Threats to internal validity are factors other than the intervention that might account for the outcomes, whereas threats to external validity are factors such as sample size and comparability of different groups.

Many different types of **bias** can affect the credibility of research. Bias will affect the results of a study in a particular direction, favouring either the treatment group or the **control group**. It is important to know in which

direction a **bias** may be influencing the results. The most common types of bias fall into one of three categories:

Sample or selection bias, which includes:

- volunteer or referral bias
- attention bias

Measurement or detection bias, which includes issues related to how the outcome of interest was measured; for example:

- number of outcome measures used
- lack of 'masked' or 'independent' evaluation
- recall or memory bias

Intervention or performance bias, which involves bias related to how the treatment itself was carried out; such bias can arise from factors such as:

- contamination
- co-intervention
- timing of intervention
- site of intervention
- different administrators of the intervention

The remainder of this section describes these types of **bias** and provides potential ways to avoid them in a research study. It is worth noting that errors related to measurement in a research study are categorised as being either random or systematic. While random error is inherently unpredictable and hence cannot be controlled (i.e. caused by sources that are not immediately obvious), systematic error (i.e. imperfect calibration of measurement instruments) is predictable and so can be identified and eliminated. Measurement error will be discussed in more detail in coming chapters. The remainder of this section describes the aforementioned types of bias and provides potential ways to avoid them in a research study.

SAMPLE OR SELECTION BIAS

Volunteer or referral bias

Description: This type of **bias** commonly occurs when the participants have volunteered to be part of the research study, perhaps via an advertisement online or in the newspaper. The problem is that the people who have volunteered may be more motivated than other people from the population.

Potential solution: Where possible, participants should be randomly selected from the population, but sometimes practical or ethical reasons make this difficult. Inviting potential participants (e.g. from treatment waiting lists at a clinic) is preferable to advertising for volunteers.

Attention bias

Description: If people are aware of the intention of the study, they might perform differently than those who are not. This is seen with the '**placebo** effect', in which the expectation of the treatment leading to improvement sometimes causes relief of symptoms.

Potential solution: A **control group** should be included, and if possible should receive a placebo treatment. It is not always possible to use a placebo, particularly when the nature of the treatment cannot be concealed from participants (e.g. if a trial is comparing treatment delivered via a webcam with treatment delivered face-to-face).

MEASUREMENT OR DETECTION BIAS

Number of outcome measures used

Description: If only one outcome measure is used, there can be **bias** in the way the measure itself evaluates the outcome. For example, when measuring complex outcomes such as improvements in activities of daily living, it may not be possible for one outcome measure to cover all aspects of the outcome of interest.

Potential solution: Researchers should use a range of outcome measures that capture the key aspects of the outcome of interest.

Lack of 'masked' or 'independent' evaluation

Description: If the researcher is aware of which group a participant is allocated to, or of which treatment they received, the researcher may influence the results in one direction or another.

Potential solution: Where possible, the researcher and assessors (i.e. those who are measuring the outcomes) should be blinded to which group participants have been allocated to. As above, this is not always possible for researchers who are delivering a behavioural treatment; however, outcomes should be measured by independent assessors who are blind to the group participants have been allocated to.

Recall or memory bias

Description: If outcomes are measured using self-report tools that require the participant to recall past events, a participant may recall only fond or positive memories.

Potential solution: Where possible, chosen outcome measures should not rely on participants having to recall past events; instead, measurement should occur in a timely manner.

INTERVENTION OR PERFORMANCE BIAS

Contamination

Description: The **control group** (who should not have received treatment) inadvertently receives treatment; thus, the difference on a particular measure may be reduced.

Potential solution: Researchers should have strict protocols for their research team regarding the delivery of treatment to participants and the management of the control group.

Co-intervention

Description: If a participant receives another additional intervention at the same time as the intervention under investigation, this may influence the results. For example, if a patient is taking medication while undergoing treatment, that medication may have a positive or a negative effect on the treatment under investigation.

Potential solution: Participants should receive information about the protocol of the study, and should be asked about any medications they may be taking or other interventions they may be undergoing, as part of the routine admission to the study process.

Timing of intervention

Description: If treatment is of a short duration, there may not have been enough time for any noticeable change to occur. If intervention is delivered over a long period of time, especially if it involves children, maturation (i.e. improvement irrespective of treatment) may occur.

Potential solution: Treatment protocols should be followed to ensure that treatment is delivered over the necessary time period. In addition, if treatment must be delivered over a long period, then a control group of participants who receive no treatment can be used to overcome the issue of maturation (i.e. participants in both groups would be expected to experience maturation, so any differences can be attributed to the treatment).

Site of intervention

Description: The place where the intervention takes place may affect the result. For example, if a study is investigating an intervention that focuses on developing play skills in kindergartens, then all kindergartens investigated should be similar in terms of the toys and activities children can participate in during the trial.

Potential solution: All treatment sites should be consistent.

Different administrators of the intervention

Description: Sometimes, different therapists are involved in delivering the treatment under investigation. This is problematic; for example, if one therapist is more dynamic and engaging than another, participants interacting with the more engaging therapist may respond better to treatment owing to the influence of the therapist.

Potential solution: Therapist involvement should be equal and consistent between all treatment groups. Therapists should be trained to deliver the treatment, and monitored for the accuracy of their delivery.

In the following video Associate Professor William McGuiness highlights the issues of **bias** in the context of wound research:

https://doi.org/10.26181/5c118fdb64ca2

Identifying other limitations in quantitative research

In addition to the common types of **bias** described above, there are other aspects of a study that you should evaluate before deciding whether the findings can inform your clinical practice. These aspects include the characteristics of the sample, the number of people who dropped out of the study, and the method and frequency of measurement, as outlined below.

Sample

Questions to ask in relation to the sample are:

- How many participants were involved, and was this a sufficient number to be able to generalise the results to the population?
- Was the sample size justified? Preliminary studies may involve small sample sizes, but if the researchers aim to demonstrate efficacy, then a larger sample is needed.
- Was a clear description of the key characteristics of the same provided (e.g. the age and gender of participants, and the onset of the health condition of interest)?
- If there were multiple groups in the study, were they comparable in terms of size and participant characteristics?
- Were appropriate inclusion and exclusion criteria described?

Drop outs (also known as experimental mortality or experimental attrition)

Questions to ask in relation to drops outs are:

- Were the number of drops outs reported?
- Were the reasons for the drop outs documented?
- How did the researchers manage the analysis of the data to deal with any missing data caused by participants dropping out?

Measurement

Questions to ask in relation to measurement are:

- How frequently were the outcomes measured; for example, were they measured before and after treatment, or were short-term and long-term follow-up data also collected?
- Did the researchers report whether the outcomes measures used are well-established as being reliable and valid?

This final consideration (regarding the **reliability** and **validity** of outcomes measures) provides a nice segue to the next chapter, in which we look at these aspects of measurement, which are often referred to as the 'psychometric properties'.

Further reading

Sedgwick, P. (n.d.). Internal and external validity. BMJ, 340(7749), 767 756-767.

Chapter 7

Measure-ment: reliability and validity

Introduction and learning outcomes

Health professionals need to be aware of the crucial role that measurement plays in clinical practice. For example, often, they will need to measure the important characteristics and abilities of their patients or clients, and make clinical decisions based on this information. Additionally, health professionals will need to measure change over time, to see whether a particular intervention is having an effect. Measurement is equally important in quantitative research. As consumers of research, we need to have confidence that researchers have collected data in such a way that we can trust that the results are accurate. If we cannot have confidence in the way in which measurements have occurred, then we certainly cannot believe the results of the research and we cannot use the research to help us make good clinical decisions.

In your clinical role and as a consumer of research you will read about, and use, many different types of measurement, including **objective** and **subjective** measures. Objective measures involve impartial measurement; for example, using a stop watch to time how long it takes someone to run 100 metres. Subjective measures are influenced by the observer's personal judgement; for example, rating the severity of a child's stutter using a scale of 0–10. Even if a rating scale with a numerical value is used to make the measurement, this is still influenced by the interpretation of the observer. You may also read about, and use, other assessment and diagnostic tests that provide health professionals with information that helps them to determine the likelihood that a person has (or does not have) a health condition. Irrespective of the purpose of a particular measure, there are two important considerations: is the measure valid and is the measure reliable? Answering these questions relates to establishing the psychometric properties of a particular measure.

KEY LEARNING OUTCOME

Critically appraise the **validity** and **reliability** of measurement reported within a journal article.

ENABLING OUTCOMES

Once you have completed this chapter, you should be able to:

- define the following terms related to reliability: systematic and random error, intra-rater and inter-rater reliability, and test–retest reliability
- define the following terms related to measurement validity: face, content, construct and criterion validity
- identify commonly used statistics that are reported, related to measurement reliability

Measurement validity

When considering measurement in a clinical role, or when reading research, we need to be able to trust that measurements are actually measuring what they are purporting to measure. It would seem obvious that using a tape measure is a valid measure of length in centimetres (although we could still confirm its accuracy if necessary), but many concepts are not as easily measured. For example, anxiety is one of numerous health constructs that can be measured using questionnaires; however, when a questionnaire is used, we need to be sure that the questions are actually measuring anxiety and not a separate (but sometimes related) construct such as depression. To do this, we need to be aware of the following types of measurement validity – face, content, criterion and construct – as discussed below.

Face validity

Face validity means the degree to which a measurement appears, on the face of it, to measure what it is supposed to. A researcher who develops a new measure of a concept should at the very least establish that the measure appears to reflect the content of the concept in question. This might be established by pre-testing the measurement instrument with colleagues or experts, who can judge whether the measure appears to reflect the concept.

Content validity

Content validity refers to the extent to which **variables** cover the entire content, or all the major dimensions, of the concept being measured. For example, an examination in a particular subject may be the only way to gauge a student's academic performance. To ensure that the test has content validity, the lecturer might write down all the material covered in the semester, and then ask a question related to each topic. You may have had an experience where, after sitting an exam, you thought it did not reflect what was covered in the course; in such a situation, you could argue that the examination measure lacked content validity.

Criterion validity

Criterion validity means that the measure is valid if scores correlate with other measures of the same concept (usually a well-accepted or 'gold standard' method of measurement). If we take the concept of **subjective** measurement of health as an example, there are established valid measures such as the SF 36 (a list of 36 questions that subjectively measures health). To establish the criterion validity of another subjective measurement of health, we would expect that scores on both measures would correlate with each other.

Construct validity

Construct validity is concerned with how well a measure conforms to theoretical expectations, or how well it measures a theoretical or abstract construct. It is argued that IQ (intelligence quotient) tests provide an example of low construct validity. Although IQ tests were developed to measure intelligence, it has been argued that they only measure one dimension of intelligence – the potential to achieve in a white, middle class academic system – and that other dimensions of intelligence remain untapped. In establishing construct validity, the researcher needs to explicitly describe the theoretical concepts and how they relate to each other.

Table 7.1 summarises these different types of validity in terms of when they are used, how they are carried out and what they mean.

Table 7.1 Different types of validity – when they are used, how they are computed and what they mean

Type of Validity	When You Use It	How You Do It	An Example of What You Can Say When You're Done
Content Validity	When you want to know whether a sample of items truly reflects an entire universe of items in a certain topic	Ask Mr. or Ms. Expert to make a judgment that the test items reflect the universe of items in the topic being measured.	My weekly quiz in my stat class fairly assesses the chapter's content.
Criterion Validity	When you want to know if test scores are systematically related to other criteria that indicate that the test taker is competent in a certain area	Correlate the scores from the test with some other measure that is already valid and that assess the same set of abilities.	The EATS test (of culinary skills) has been shown to be correlated with being a fine chef 2 years after culinary school (an example of predictive validity)
Construct Validity	When you want to know if a test measures some underlying psychological construct	Correlate the set of test scores with some theorised outcome that reflects the construct for the test is being designed.	It's true—men who participate in body contact a physically dangerous sports score higher on the TEST (osterone) test of aggression.

Adapted from figure "Types of Validity" by SAGE College; used with permission

Measurement reliability

In health practice, **reliability** is the extent to which a measurement instrument or performance is dependable, stable and consistent when assessed under identical conditions. Few, if any, of the measurement instruments used in health care are 100% reliable. Thus, most measurement instruments possess some 'measurement error'; this is to be expected, given

that humans are inherently fallible when taking measurements, which adds to the measurement error. The greater the measurement error, the lower the reliability.

When you are reviewing evidence from journal articles, you need to make decisions about whether reported measurements are reliable. If they are not reliable, this will affect the accuracy of the study findings, and therefore mean that you are less likely to adopt this evidence in your health practice.

When considering whether you can adopt the findings of an article, or generalise the results to your clients or workplace, remember that reliability is 'estimated' from a sample that is representative of a specific population. You therefore need to keep in mind that the true reliability will differ slightly in the population of interest.

Reliability can be broken down into several layers, as discussed below. The research articles that you review will generally report reliability in terms of consistency and or agreement, which are discussed below.

CONSISTENCY

Most studies reporting the reliability of an observation report consistency using a correlation statistic that represents the strength of the association between two measurements. The correct term for a correlation statistic is 'correlation coefficient'. Commonly reported correlation coefficients include Pearson's *r* and the intraclass correlation coefficient (ICC). Both of these coefficients have values ranging from 0 to 1, where 0 indicates 'no correlation' and 1 indicates 'perfect correlation'. These correlation statistics are described in more detail below.

AGREEMENT

Ideally, the studies you review will not only report the correlation between two sets of observations (e.g. r value or ICC), but also the level of agreement. For example, on average, does measurement A differ from measurement B by 10 cm, or 5 °C or 30 minutes? Understanding the level of agreement for a measurement is useful in health practice because it allows us to make decisions about whether the magnitude or size of the error is acceptable or unacceptable.

Consistency and agreement can be affected by multiple sources of error. For example, consider a study reporting the reliability of measuring core temperature using a handheld ear thermometer, with measurements taken two minutes apart. Many factors will influence the tester's ability to obtain a reliable result, including the ability of the tester to correctly insert the thermometer into the earhole, the stability of the air temperature in the room and the stability of the participant's core temperature. Some of these

factors can be controlled, allowing us to determine the reliability of the tester and the device itself. With this mind, reliability in health research is generally divided into different sources: reliability of the test as a whole (i.e. test–retest reliability), reliability of the tester (i.e. intra-tester reliability) and reliability of different testers (i.e. inter-tester reliability). These sources of reliability are discussed below.

Test–retest reliability

Test–retest reliability evaluates the stability of a measurement obtained on two different occasions when we would expect no change in the construct being measured. It reflects whether a certain task can render reliable results or whether it is highly dependent on the situation, or on the state of the subject (i.e. when presenting the same task to the same subjects two or more times). You would expect, for example, that because IQ is a relatively stable construct, if you were to have your IQ measured by a psychologist tomorrow and then again in six months' time, a measure of IQ that had good test–retest reliability would give you a similar IQ on both occasions.

EXAMPLE 7.1 RANDOM BREATH-TESTING OF DRIVERS TO ESTIMATE BLOOD ALCOHOL LEVELS

You may have been breath-tested or seen your parents blow into a hand-held breath-testing device. Imagine you were asked to investigate the test–retest reliability of this device. Think about what sort of experiment you could do to determine whether the procedure is reliable in estimating blood alcohol levels.

One experiment you could undertake, would be to observe a random roadside breath testing station. The police could seek permission from drivers to repeat the test after a five minute period under precisely the same conditions to ensure that the test is stable (i.e. you are re-measuring the same characteristic and assuming that the characteristic being measured is stable).

The scatterplot in Figure 7.1 shows the test and re-test values for each person for measurement 1 and measurement 2. Each dot represents measurement 1 (vertical axis) and measurement 2 (horizontal axis) for one of the friends at your party.

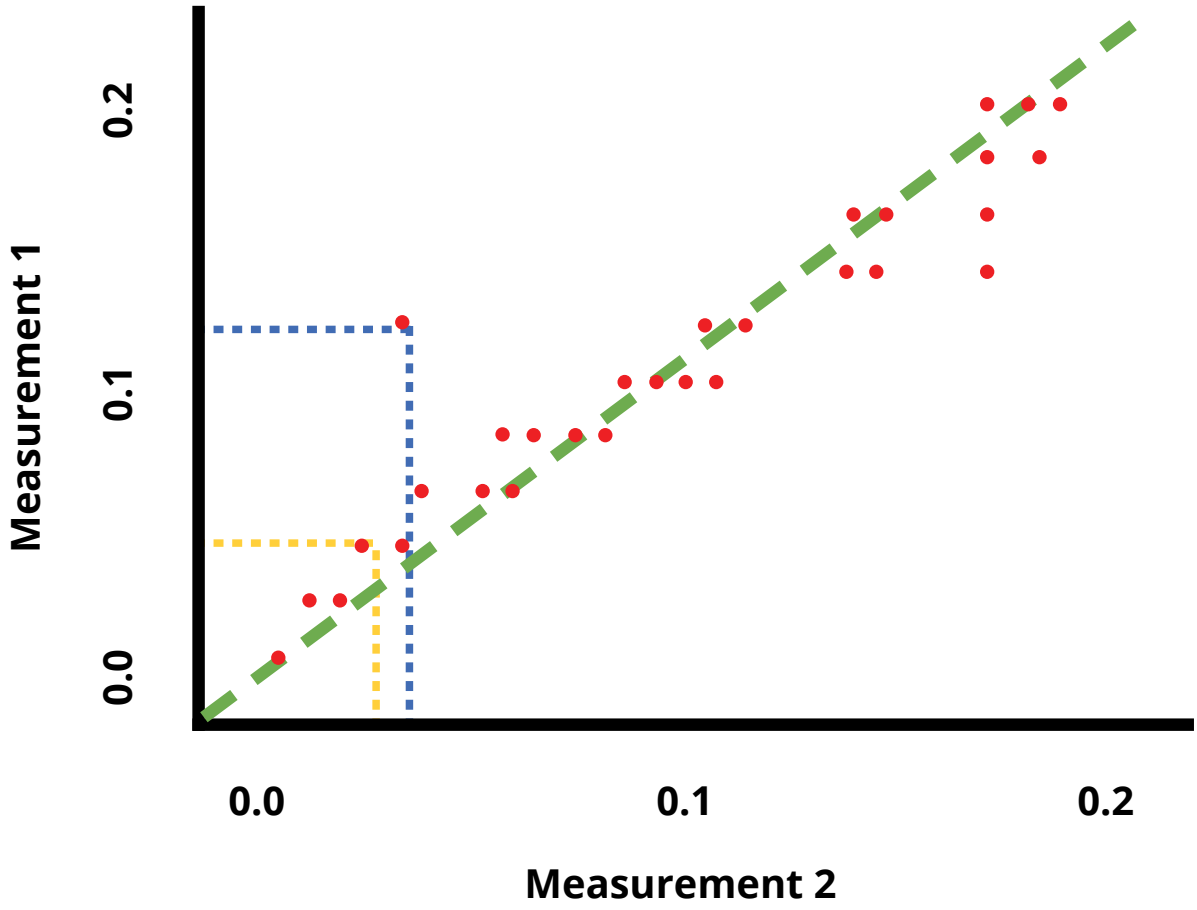
Can you see that the friend represented by the red dotted lines has about the same blood alcohol reading at measurement 1 and measurement 2? In contrast, the friend represented by the blue dotted lines has quite

different blood alcohol levels detected at measurement 1 and measurement 2 (with measurement 2 being a higher value)?

If the **reliability** for this study was 'perfect', all of the friends' blood alcohol results for measurements 1 and 2 would sit on the green line, giving a reliability coefficient of 1.0.

This experiment would give you the test–retest reliability of the blood alcohol breath test. In this example, test–retest reliability is being performed for the test as a whole; that is, you are not specifically determining the reliability of the tester or the reliability of different testers.

Figure 7.1 Scatterplot



Intra-rater reliability

Intra-rater reliability evaluates the ability of a single rater to obtain the same result when presented repeatedly with the same observations. For intra-rater reliability, the observation, object or performance being tested is fixed, which means that any variation between measurements can be attributed to the tester.

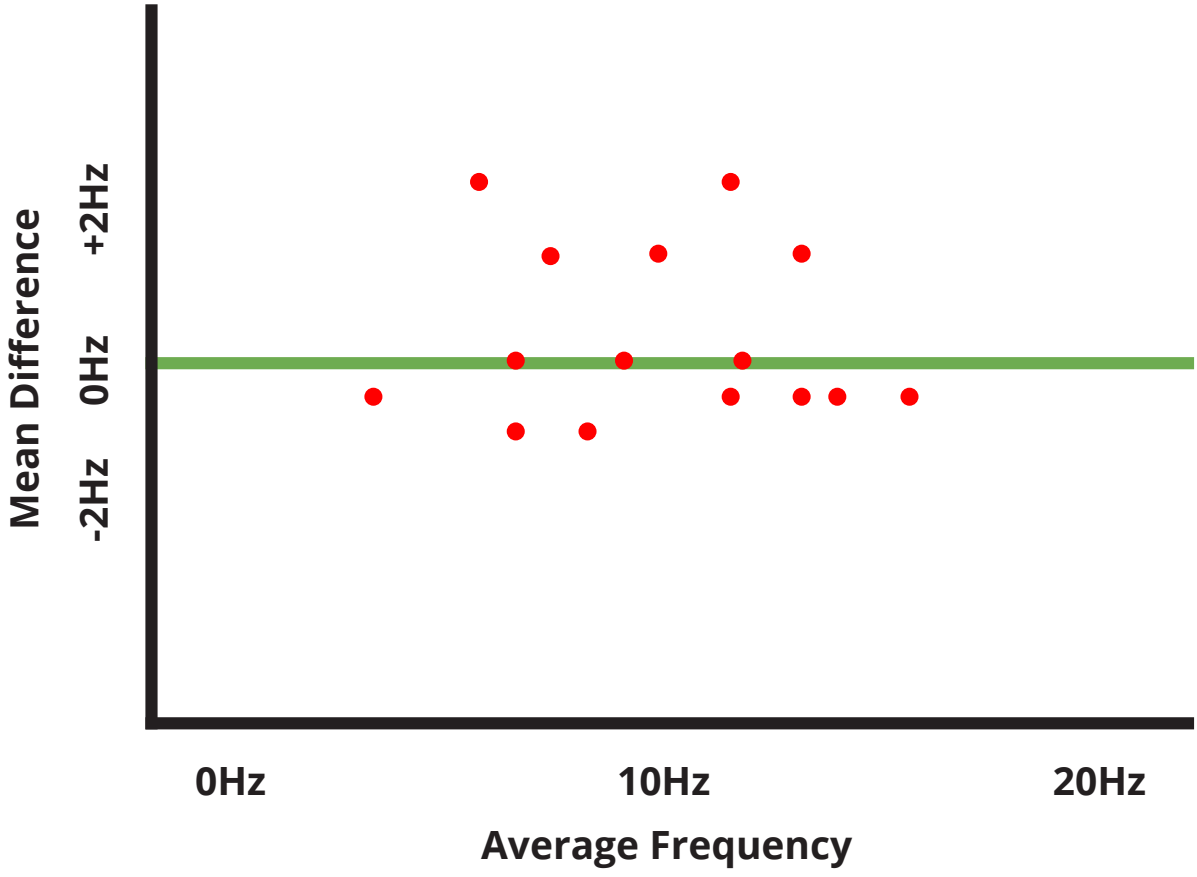
EXAMPLE 7.2 ANALYSIS OF PRINTED VOICE SIGNAL FREQUENCIES TO DETERMINE INTRA-TESTER RELIABILITY

Imagine that you are determining your **reliability** at analysing printed voice signal frequencies. To do this, you analyse printed voice signal

frequencies from 15 children with a voice disorder. You analyse each printout on Day 1, and then re-analyse the same printouts on Day 10 to determine intra-tester reliability.

The findings from your reliability study on voice signal frequencies are presented in Figure 7.2 on a Bland-Altman reliability plot. This type of plot is an informative way of presenting the results, with the magnitude of error on the vertical axis and the average measurement for each observation or individual on the horizontal axis. Each dot indicates the 'difference' between the measurements on Day 1 and Day 10 for each individual participant. For three measurements, the difference between the Day 1 and Day 10 values was zero.

Figure 7.2 Mean difference and average frequency (intra-tester reliability)



Inter-rater reliability

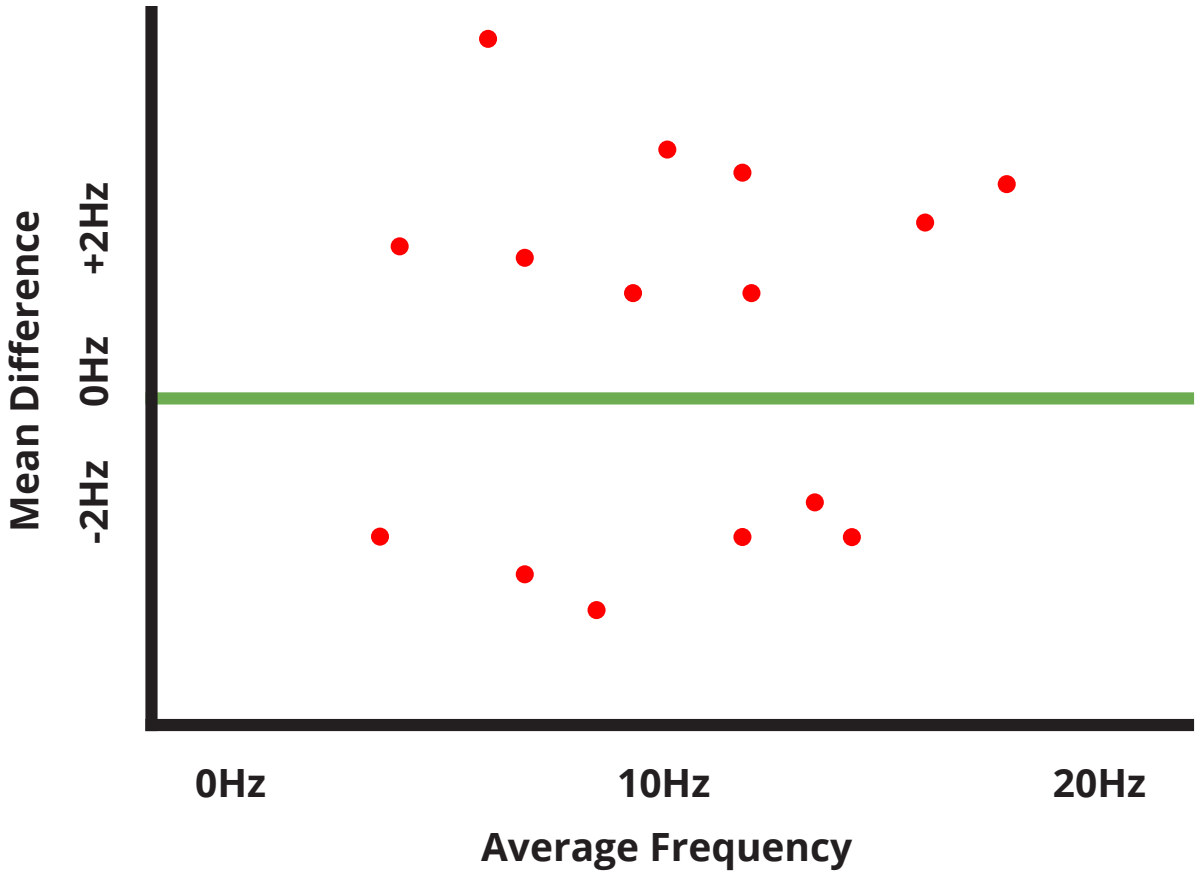
Inter-rater reliability evaluates the ability of different raters to obtain the same measurement relative to each other. As with intra-rater reliability, the observation, object or performance being measured is fixed or well controlled, so that any variation between measurements can be attributed to the different raters.

EXAMPLE 7.3 ANALYSIS OF PRINTED VOICE SIGNAL FREQUENCIES TO DETERMINE INTER-RATER RELIABILITY

Let us consider the example of the evaluation of the printed voice signal frequencies given in Example 7.2: however, this time we are interested in the inter-rater reliability. The sample printout is assessed by two independent raters and their scores are later compared.

The findings from the study are presented in Figure 7.3 on a Bland-Altman reliability plot, which shows the differences between the raters. This time, there are no scores with a zero mean difference, and there is less 'agreement' in scores compared with the Bland-Altman plot for intra-tester reliability above. This shows that differences between people are greater than differences in the performance of one person.

Figure 7.3 Mean differences and average frequency (inter-tester reliability)



Random and systematic error

As noted previously, there are many sources of 'error' in research that reduce the **reliability** of measurement. Error can be broadly classified as *random* or *systematic*. Random error is unpredictable and is scattered about the true value – like darts around the bullseye on a dartboard. Systematic error, however, is usually predictable and therefore directional.

A classic example of systematic error in health-care research is the improvement in human performance with repeated testing. Imagine that you are investigating the reliability of performing a hamstring stretch with repeated measurements every two minutes. You would expect flexibility to improve with each stretch, you might also expect that some participant learning would occur (i.e. participants would improve their proficiency in performing the stretch).

'Stretching Fitness' from flickr used under CC BY 2.0

If you carried out four tests, you would probably find that the later tests (i.e. tests 3 and 4) would have a higher value than the earlier tests (i.e. tests 1 and 2). You might interpret these findings as indicating that the hamstring flexibility test is not reliable, because the values are different; however, what is also evident is a systematic increase in values over time.

Conversely, random error in measurement is due to unpredictable factors such as tester fatigue, inattention or just simple mistakes. For example, in a research study where measurements of a participants' height are taken, a tester might make inaccurate measurements due to inadvertently stretching the tape measure more on some occasions than others. This random error leads to inconsistency in the measurements.

Correlation coefficients, their confidence interval and p value

Some research articles report absolute differences (or level of agreement) between any two measurements, whereas other studies only report a correlation coefficient such as the r or ICC statistics. These coefficients have

values ranging from 0 to 1, where 0 indicates 'no correlation' and 1 indicates 'perfect correlation'; however, sometimes a negative value is reported (i.e. ranging from 0 to –1), indicating a negative correlation. The following guide could be used to interpret correlations, however note that these are not definitive interpretations:

- correlations from 0 to 0.25 (or –0.25) indicate no relationship (or poor **reliability**)
- those from 0.25 to 0.50 (–0.25 to –0.50) indicate a fair relationship or fair reliability
- those from 0.50 to 0.75 (–0.50 to –0.75) indicate a moderate to good relationship or moderate to good reliability
- those greater than 0.75 (or –0.75) indicate very good to excellent relationship or reliability

In addition to correlation coefficients, there are two more important statistics that you need to consider – the confidence interval (CI) and the **p-value**. The CI is a range or margin of error for the correlation coefficient; it gives an estimate of the true **reliability** of the measurement being performed. You can expect the true reliability to have a value somewhere within the range of the CI. The p-value indicates the level of statistical significance. Usually, when p is less than 0.05 (i.e. $p<0.05$) we conclude that a finding is 'statistically significant', and when p is greater than 0.05 (i.e. $p>0.05$) we conclude that statistical significance was not detected.

A more thorough explanation about the meaning of p-values and confidence intervals are covered in chapters 10 and 11. They are briefly described here to help place them into the context of correlation coefficients.

EXAMPLE 7.4 ASSESSMENT OF SYSTOLIC BLOOD PRESSURE

Imagine that you review a research article that reports the intra-rater reliability for an occupational therapist assessing systolic blood pressure. The ICC is reported as 0.70, indicating that the intra-rater reliability of the occupational therapist is 'moderate' to 'good'. However, imagine that the research article also reported a 95% CI of 0.49–0.91 for the ICC. This would mean that there is a 95% chance that the true ICC is actually somewhere between 0.49 and 0.91, which in turn would mean that the true reliability might only be 'fair'. You can see that it is important to pay particular attention to the lower limit of the CI when interpreting the ICC, because you want to know how much measurement error is present. If the measurement error is too great, you cannot have confidence in the finding or in using that measurement in your practice.

Now imagine that the research article reported a p-value of 0.015 for the ICC. Because this value is less than 0.05, we can say that the ICC of 0.70 was statistically significant (i.e. $p<0.05$).

Further reading

Measurement (2013) p.105. In. S. Polgar & S.A. Thomas, Introduction to research in the health sciences, 6th Ed., Churchill Livingstone Elsevier.

Polit, D. (2015). Assessing measurement in health: Beyond reliability and validity. International Journal of Nursing Studies, 52(11), 1746-1753.

Chapter 8

Measurement and analysis: outcome measures and scales of measurement in health research

Introduction and learning outcomes

In the following video Simon Pampena discusses some of the basics of statistics:

https://doi.org/10.26181/5c11e63fb6f84

When working as a health professional and interpreting health research (i.e. evidence), you will be required to summarise and communicate statistics. Some examples of how you might use these skills in the workplace include:

analysing data from a database of clients with diabetes in a hospital

preparing effective marketing campaigns for your private practice

reviewing a piece of literature to report to a journal club in the workplace

making decisions about the suitability of a new treatment for your clients.

These are just some of the many situations where you will need to understand how to organise and present different types of data.

KEY LEARNING OUTCOMES

- Describe and categorise outcome measurements in the context of a research article.
- Outline **methods** for organising or representing, and summarising data

ENABLING OUTCOMES

Once you have completed this chapter, you should be able to:

- describe the various scales of measurement
- organise and present different types of data: nominal, ordinal, interval and ratio

Types of measurement scale

In the different scenarios listed in Section 8.1, the data you obtain will have distinct measurement properties. For example, in the scenario of analysing a client database in a hospital, your analysis might involve auditing clients' ethnic background, age group and gender; the duration of time since they were diagnosed with diabetes; the clients' peak core body temperature at the time they were admitted to hospital; or how satisfied the clients were with the service provided by your clinic. These various types of data are presented in different units of measurements; hence, they must be summarised and analysed differently. They can be categorised into four measurement types – nominal, ordinal, interval and ratio – each of which is discussed below.

NOMINAL SCALES

A nominal scale is the simplest type of measurement scale, because the data or variable is allocated a name or identification number, and is categorised. Nominal data are sometimes referred to as 'discrete' or 'categorical' data. In the example above, each client's ethnic background is a nominal scaled variable (e.g. American, Chinese or Irish). Often, the **variables** are coded numerically (e.g. country 1 or country 2), but there is no relationship between the allocated numbers (i.e. 2 is not greater than 1 in this instance, it just represents a different code).

ORDINAL SCALES

An ordinal scale involves the rank ordering of a variable. As with nominal data, ordinal data are sometimes referred to as 'discrete' or 'categorical' data. Ordinal measurements describe order, but not relative size or degree of difference between the items measured. For example, in your audit of the hospital database, you might assign clients into the following age groups: Group 1: 36–40, Group 2: 41–45, Group 3: 46–50 and Group 4: 51–55 years. The group numbers are assigned on an ordinal scale to signify order or rank.

INTERVAL SCALES

Data presented on an interval scale has no absolute zero. Interval-scaled data are also referred to as 'continuous' data. For example, in your audit of the hospital database, you might record each client's peak core body temperature when the patient was admitted to hospital. Temperature (in degrees Celsius) is an interval-scaled variable because it has no absolute zero (i.e. 0 °C represents 'freezing point', not the point at which there is no heat). **Variables** measured on the interval scale are called *interval variables* or *scaled variables* because they have units of measurement. Ratios between numbers on the scale are not meaningful, so operations such as multiplication and division cannot be carried out directly. However, ratios of differences can be expressed; for example, one difference can be twice as big as another.

RATIO SCALES

The ratio scale involves all the characteristics of the other scales, but also has the characteristic of an absolute zero. As with interval scales, ratio-scaled data are referred to as 'continuous' data. For example, in your audit of the hospital database, you might record the duration of time since each client was diagnosed with diabetes. Presuming that all of the clients had diabetes, none of them could have had diabetes for less than zero days.

One of the powerful characteristics of ratio-scaled data is that it is 'backwards compatible'; that is, it can be re-scaled into nominal, ordinal and ratio scales. An example of this is shown in Table 8.1.

Table 8.1 Levels of measurement

Characteristics of levels of measurement	Nominal	Ordinal	Interval	Ratio
Qualitative, descriptive	Y	N	N	N
Rank-ordered according to place magnitude	N	Y	N	N
Equal intervals with no absolute zero	N	N	Y	N
Absolute zero	N	N	N	Y
Examples of variables (types of data)				
Preferred AFL team (Collingwood, Carlton, Melbourne)	Y	N	N	N
Student feedback on subject (Overall, how satisfied are you with the subject: very satisfied, unsatisfied, neutral, satisfied, very satisfied)	N	Y	N	N

Table 8.1 Levels of measurement (continued)

Characteristics of levels of measurement	Nominal	Ordinal	Interval	Ratio
Examples of variables (types of data)				
Range of movement of the hip joint (negative values = extension; positive values=flexion)	N	N	Y	N
Heart rate measured during your exam (220 BPM)	N	N	N	Y
Terminology				
Also referred to as…	Categorical data Discontinuous data Discrete data	Categorical data Discontinuous data Discrete data	Continuous	Continuous
Example of how ratio scaled data can be transformed to low-order measurement scales.	After knee surgery subjects were allocated to either a 'no knee pain group' or a 'knee pain group'	After knee surgery subjects were allocated to one of the following groups: 1. low pain (0-25); mild pain (26-50); moderate pain (51-75) and severe pain (76-100)	After knee surgery subjects had knee pain scores ranging from -50 (reduced knee pain) to +10 (increased knee pain)	After knee surgery subjects had knee pain scores ranging from 0 (no pain) to +100 (severe pain)

Summarising different measurement scales

Once you have collected data from a source (e.g. a client database, a research experiment or some type of marketing database), you eventually need to undertake some sort of statistical analysis. Before you can start a statistical analysis, you need to know which method to use to summarise the type of data you have obtained. Different methods are used for summarising nominal, ordinal, interval and ratio-scaled data. For example, you cannot calculate the 'mean' from a frequency count of 10 males and 27 females – this situation is explained further in Chapter 10.

EXAMPLE 8.1 SOCIAL NETWORKING SCENARIO

As part of a marketing campaign, your employer has asked you to compare demographic and usage characteristics of Australian users of the social networking sites Facebook and Instagram (your employer has paid a lot of money to access these databases). Your employer is particularly interested

in the following characteristics from users aged 25–35 years:

A. How many different embedded games they play on Facebook and Instagram (e.g. Words with Friends and Scrabble).

B. How many minutes they use these platforms per day, expressed as low use (0–20 minutes), moderate use (>20 to <60 minutes) and high use (>60 minutes).

C. Occupation type.

D. On average, how often they post messages (e.g. comments and status updates) relative to the 'world average' of 10 per day.

As a first step, can you correctly assign a scale of measurement (i.e. nominal, ordinal, interval or ratio) to the usage characteristics (A–D) given above?

DISCONTINUOUS DATA

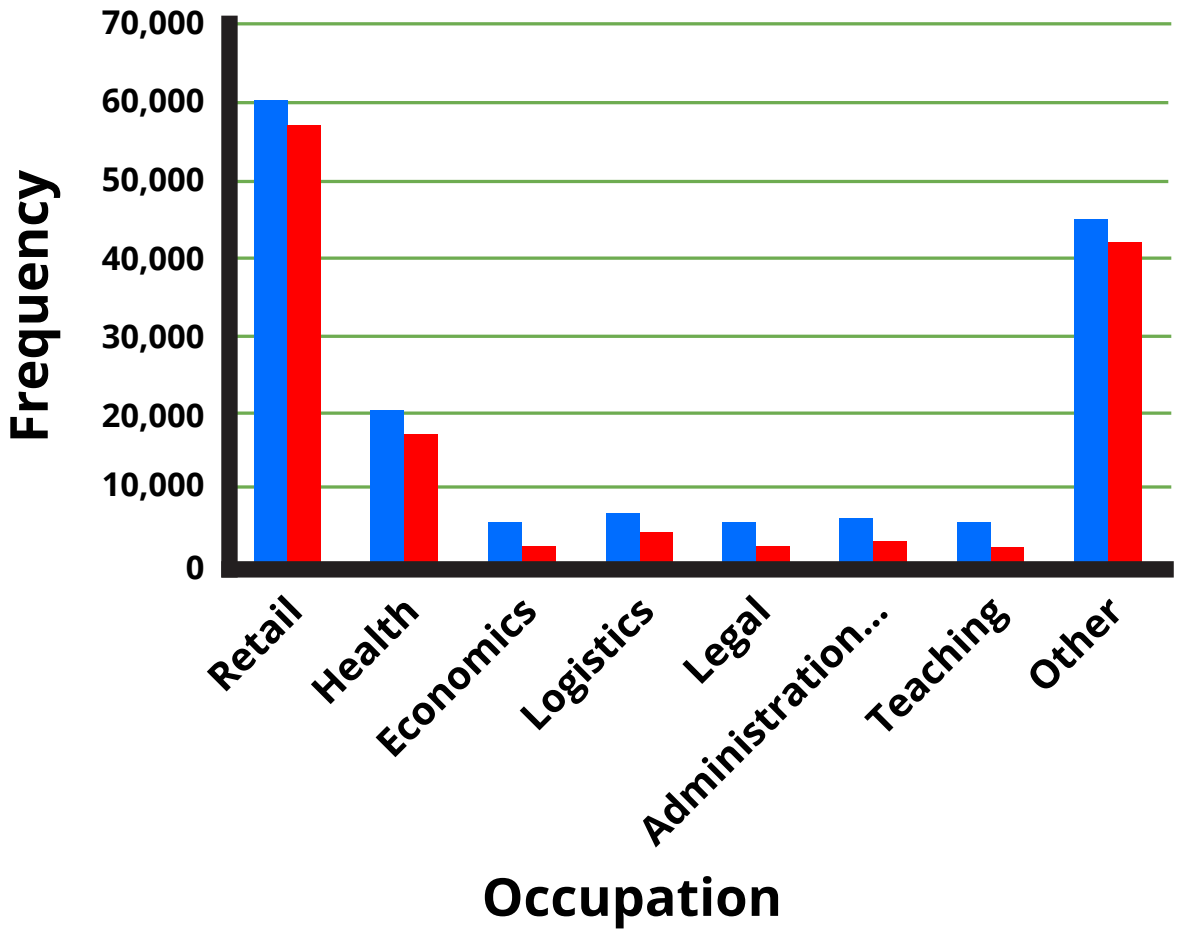
Nominal-scaled data

Hopefully, you selected item C (i.e. occupation type) as nominal-scaled data. Summarising nominal data usually involves counting the number of cases that fall into each category. In your analysis of the social network data, you might form the categories shown in Table 8.2 and Figure 8.1, and present the data as a frequency (*f*) distribution.

Table 8.2 Frequency distribution social networking scenario

Occupation	Facebook	Instagram
	f	*f*
Retail	60,150	57,180
Health	20,027	17,057
Economics	5,303	2,333
Logistics	7,005	4,035
Legal	5,545	2,575
Administration	5,951	2,981
Teaching	5,502	2,532
Other	45,124	42,154
Total	*N = 154,607*	*N = 130,847*

Figure 8.1 Occupation of Australian users of **Facebook** and **Instagram**



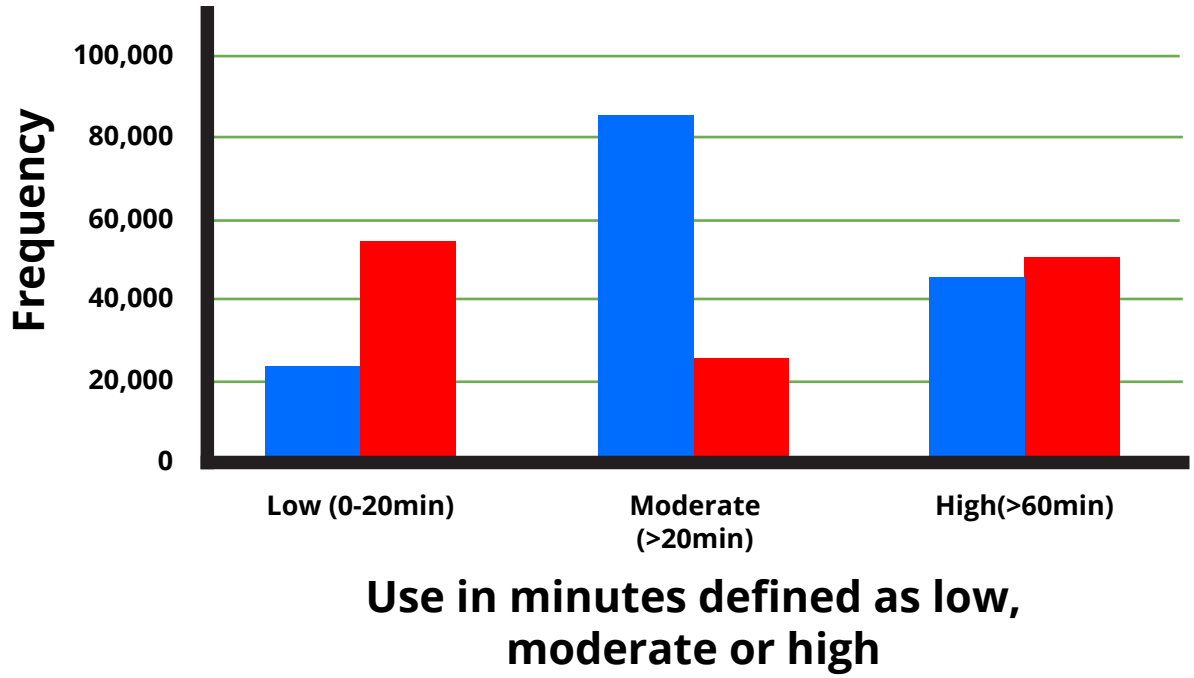
Ordinal-scaled data

Hopefully, you selected item B (i.e. minutes of use per day) as ordinal-scaled data. Summarising ordinal data usually involves counting the number of cases that fall into each category. In your analysis of the social network data you might form the categories shown in Table 8.3 and Figure 8.2, and present the data as frequency (*f*) distribution.

Table 8.3 Use per day social networking scenario

Use per day	Facebook	Instagram
Low: 0-20min	23,456	54,321
Moderate: >20min, >60min	85,499	25,641
High: >60min	45,652	50,885
Total	*N = 154,607*	*N = 130,847*

Figure 8.2 Use per day of **Facebook** and **Instagram**



Continuous data

Hopefully, you selected item D (i.e. frequency of posting messages relative to the world average) as interval-scaled data. Viewing raw interval-scaled data (especially for >280,000 individual social network users!) is cumbersome and uninformative; hence, the data need to be summarised so that they are easier to understand. There are several approaches to summarising interval-scaled data. One method involves presenting the observations as grouped frequency distributions, as shown in Table 8.4 and Figure 8.3.

Table 8.4 Frequency of posts in social networking scenario

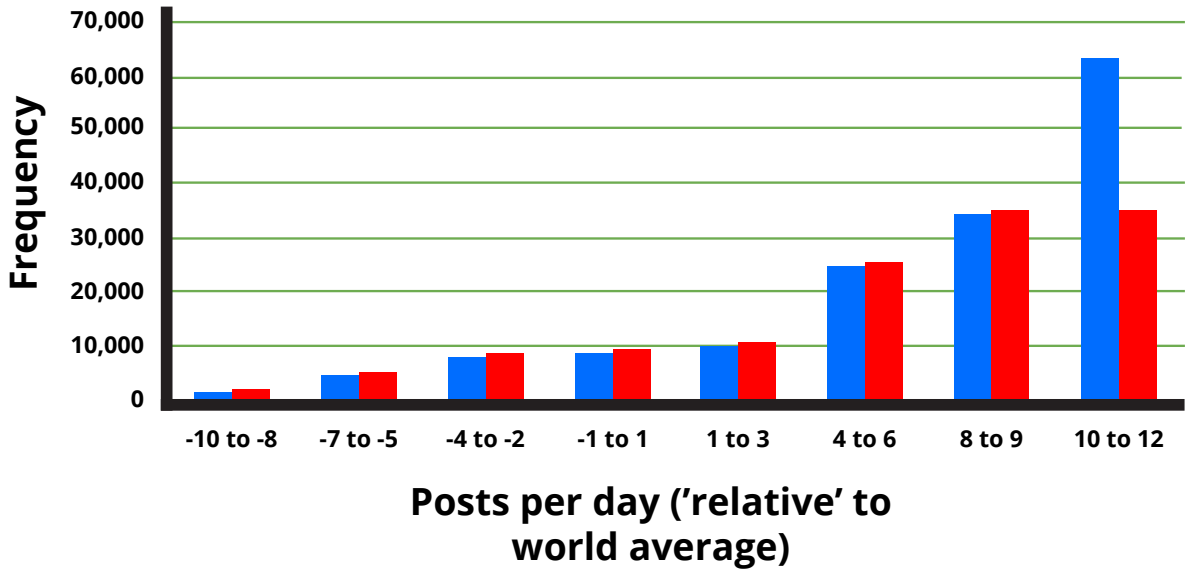
Frequency of posts per day ('relative' to world average)	Facebook	Instagram
	f	*f*
(-10 to -8)	1,202	1,789
(-7 to -5)	4,563	5,150
(-4 to -2)	7,838	8,425
(-1 to 1)	8,741	9,328
(1 to 3)	9,851	10,438
(4 to 6)	24,876	25,463
(8 to 9)	34,502	25,089

Table 8.4 Frequency of posts in social networking scenario (continued)

Frequency of posts per day ('relative' to world average)	Facebook	Instagram
(10 to 12)	63,034	35,165
Total	*N = 154,607*	*N = 130,847*

You can see by the grouped frequency distribution in Figure 8.3 that the least number of messages or posts was zero, indicated by –10 (i.e. 10 less than the relative world average). You can also see that the greatest number of posts was 22, indicated by 12 (i.e. 12 more than the world average of 10).

Figure 8.3 Comments and posts per day of Australian users of **Facebook** and **Instagram**

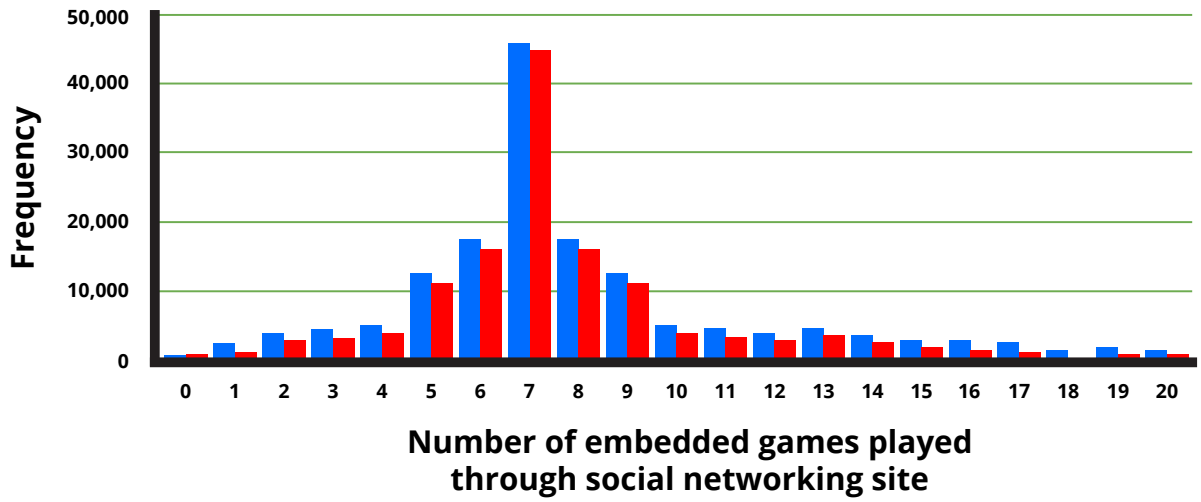


Hopefully you selected item A (i.e. number of different embedded games played) as ratio-scaled data. There are several approaches to summarising ratio-scaled data. As with the ordinal-scaled data, one method involves presenting the observations as a frequency (*f*) distribution, as shown in Table 8.5 and Figure 8.4.

Table 8.5 Number of games played per day for Australian users of Facebook and Instagram

Number of embedded games played on social networking sites	Facebook	Instagram
	f	*f*
0	601	656
1	2,282	1,094
2	3,919	2,731
3	4,371	3,183
4	4,926	3,738
5	12,438	11,250
6	17,251	16,063
7	45,788	44,600
8	17,251	11,250
9	12,438	11,250
10	4,926	3,738
11	4,371	3,183
12	3,919	2,731
13	4,532	3,344
14	3,478	2,290
15	2,794	1,606
16	2,689	1,501
17	2,332	1,144
18	1,369	181
19	1,689	501
20	1,243	656
Total	*N = 154,607*	*N = 130,847*

Figure 8.4 Number of games played per day on **Facebook** and **Instagram**



Frequency histograms and frequency polygons

In the following video Professor Geoff Cumming discusses frequency distributions:

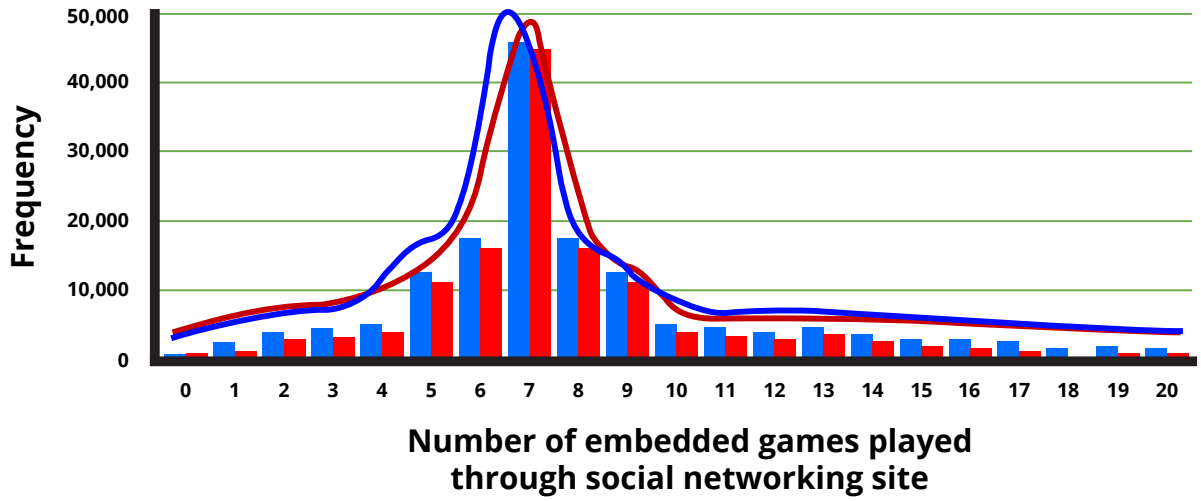
https://doi.org/10.26181/5c118fd2a2972

Figure 8.5 is a frequency histogram, with a 'point and line' plotted over the midpoint of each class of data (i.e. 0, 1, 2, 3 and so on). The point and plot line is called a 'frequency polygon', and it has two main purposes:

It is possible to **'interpolate' or estimate the frequency of missing values**. For example, from the blue and red curves in Figure 11.6 (representing Facebook and Instagram, respectively), you could easily predict the frequency or number of individuals who would play 21 or 22 games per day. Interpolation can be undertaken for continuous data but not for discrete data.

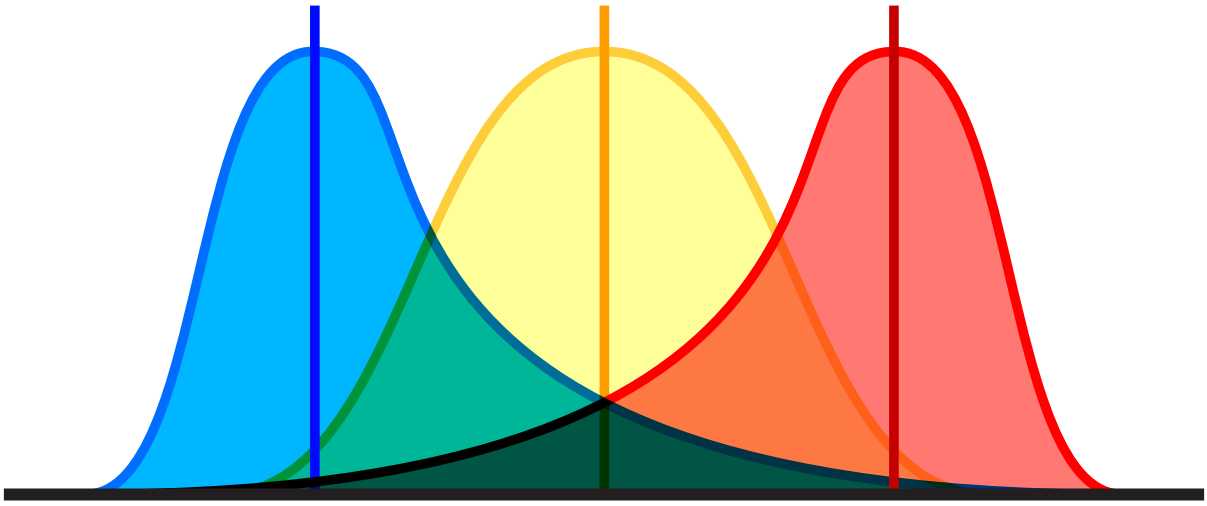
It is possible to **characterise the shape of the distribution**. The rest of this section focuses on this point, which is fundamental to your understanding of topics such as 'central tendency' and 'normality'.

Figure 8.5 Number of embedded games played through **Facebook** and **Instagram**



Measures related to human behaviour commonly exhibit one of the three 'shapes' in a frequency distribution or frequency polygon, as shown in Figure 8.6.

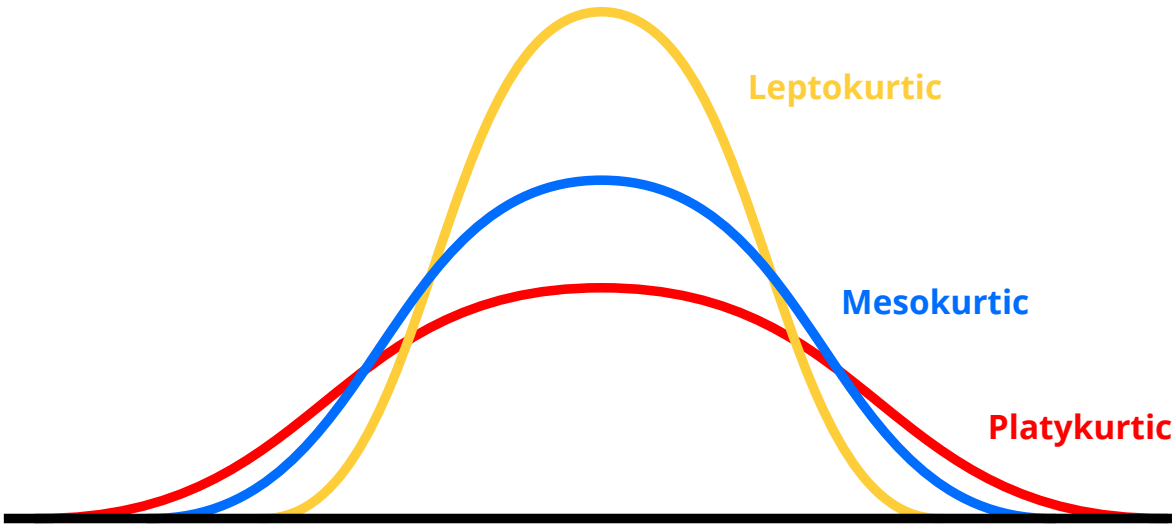
Figure 8.6 Frequency distribution or frequency polygon



1. Positively skewed (Skewed to the right) **2. Symmetrical (Bell-shaped curve)** **3. Negatively skewed (Skewed to the left)**

The vertical line in each frequency polygon in Figure 8.7 shows where the data are clustered, or where the most frequent observations are located. When the polygon shape is positively skewed, the data are clustered towards lower scores. Conversely, when the polygon shape is negatively skewed, the data are clustered towards the higher scores. In addition to characterising a data set by its skewness you also need to be aware of kurtosis. Kurtosis refers to the shape (or heaviness) of the tails of the data set. The 'heaviness' or 'lightness' in the tails is an indication of the amount of kurtosis in the data set. A normal distribution of data presents with a kurtosis of 3 (mesokurtic). If the kurtosis is less than 3 (platykurtic) the distribution will have shorter and thinner tails than normal distribution. If the kurtosis is greater than 3 (leptykurtic) the distribution will have longer and fatter tails than normal distribution.

Figure 8.7 Kurtosis



Finally, when the polygon shape is symmetrical (i.e. evenly spread on the left and right side of the graph), it indicates that the data are clustered in the centre of the distribution (also referred to as a 'normal' or 'bell-shaped' distribution). By now, you should be familiar with the characteristics of the four different types of data (nominal, ordinal, interval and ratio), and how they can be summarised in both table and graphic presentations. You should also be familiar with frequency polygons and the three common shapes.

Further reading

Measurement (2013) p.105. In. S. Polgar & S.A. Thomas, Introduction to research in the health sciences, 6th Ed., Churchill Livingstone Elsevier.

Chapter 9

Introduction to descriptive statistics and statistical inference

Introduction and learning outcomes

Chapter 8 focused on measurement scales and how these are summarised in tables, and on the use of frequency histograms and polygons. The next step is to think about how you might compare data from different categories; for example, how you might compare a Facebook dataset with an Instagram dataset. A simple way to make such a comparison is to use some basic descriptive statistics; for example, ratio, proportion, percentage, rate, mean, median, mode, standard deviation, range and CI.

At this point you might be wondering whether you have accidently enrolled in a maths degree rather than something from the health sciences! We assure you that you will need to use descriptive statistics in your role in the health workforce.

KEY LEARNING OUTCOMES

- Summarise continuous-scaled data, recognising normal and skewed distributions.
- Identify sample statistics, population statistics, descriptive statistics and inferential statistics in the context of a journal article.

ENABLING OUTCOMES

Once you have completed this chapter, you should be able to:

- define ratio, proportion, percentage, rate, mean, median, mode, standard deviation, range and CI
- define sample statistics and population statistics, descriptive statistics and inferential statistics
- identify normally and non-normally distributed data

Ratio, proportion, percentage and rate

RATIO, PROPORTION AND PERCENTAGE

The *ratio* expresses the relative frequency of one dataset relative to another (e.g. the ratio of nurses in the health sector relative to doctors). The *proportion* expresses the relative frequency of one dataset as a fraction of the whole dataset (e.g. doctors in Victoria as a proportion of all doctors employed in the

health sector in Australia). The *percentage* is the portion of a whole expressed in hundredths (e.g. the proportion of doctors in Victoria multiplied by 100). Example 9.1 illustrates these statistics.

EXAMPLE 9.1 RATIO, PROPORTION AND PERCENTAGE IN THE HEALTH SECTOR

Imagine that the number of nurses in the Victorian health sector is 105,064, and the number of doctors is 28,030. To determine the ratio, we divide the number of nurses by the number of doctors:

105,064 / 28,030 = 3.75

Therefore, the *ratio* of nurses to doctors in Victoria would be 3.75:1.

This means that, for every doctor working in the Victorian health sector, there are 3.75 nurses.

As above, imagine that the total number of doctors in Victoria 28,030. If we also imagine that there are 97,150 doctors in Australia as a whole, then we can calculate the proportion of doctors in Victoria. To determine the proportion, we divide the number of doctors in Victoria by the total number of doctors in Australia:

28,030/97,150 = 0.288524

Thus, the *proportion* of doctors in Victoria would be 0.288524.

To convert the proportion to a percentage, we multiply it by 100:

0. 288524 × 100 = 28.85%

Thus, the *percentage* of Australian doctors working in the Victorian health sector would be 28.85%.

RATES

Rates are similar to ratios and proportions; they are used to quantify the level at which a health disorder or disease is present in a given population (usually over a one-year period). You will come across two types of rates: the 'incidence' rate and the 'prevalence' rate. Example 12.2 illustrates incidence and prevalence rates using melanoma – a malignant form of skin cancer that much of Australia's population is exposed to or at risk of.

EXAMPLE 9.2 RATES OF MELANOMA

Incidence rate

The incidence rate is the number of new cases divided by the total population at risk of developing the disorder. In 2018, the estimate of new cases of melanoma was about 11,000.

For a population of about 25 million in Australia, the incidence rate = 11,000 / 25,000,000 = 0.00044

If we multiple that incidence rate (i.e. 0.00044) by 100,000, we can say that: 'In 2018, there would have been 44 new cases of skin melanoma for every 100,000 Australians'.

Prevalence rate

The prevalence rate is the number of existing cases divided by the total population at risk of developing the disorder. In 2018, the estimate of existing cases of melanoma was about 30,000.

For a population of about 25 million, the prevalence rate = 30,000 / 25,000,000 = 0.0012

If we multiple that prevalence rate (0.0012) by 100,000, we can say that: 'In 2018, there would have been about 120 individuals affected by skin melanoma for every 100,000 Australians'.

Measures of central tendency and variability

In quantitative research and in everyday life, we often use 'averages' from a set of numbers or data. Below are some examples. Table 9.1 shows sample data on the average age at which adults begin to develop arthritis in their knee joint. Table 9.2 shows sample data on the average cost of renting a shared three-bedroom house with other students.

Table 9.1 Sample data on age at development of arthritis in the knee joint

Person	Age (in years) at which arthritis first developed in knee joint
Person 1	35
Person 2	36
Person 3	25
Person 4	42
Person 5	48
Person 6	40
Person 7	32
Person 8	35
Person 9	37

Table 9.1 Sample data on age at development of arthritis in the knee joint (continued)

Person	Age (in years) at which arthritis first developed in knee joint
Person 10	46
Person 11	34
Average or mean	**37**

Table 9.2 Sample data on cost of renting in a shared 3-bedroom house

House and location	Cost per week
House 1 - Bundoora	$105.00
House 2 - Bulleen	$95.00
House 3 - Inavhoe	$145.00
House 4 - Bendigo	$100.00
House 5 - Shepparton	$95.00
House 6 - Fairfield	$160.00
House 7 - Thornbury	$107.00
House 8 - Eaglemont	$175.00
House 9 - Eaglehawk	$100.00
House 10 - Heidelberg	$100.00
House 11 - Kingsbury	$115.00
Average or mean	**$118.00**

In these examples, the average (also referred to as the 'mean') is of interest because it tells us where the data are centred or clustered. In these examples, the mean indicates that, on 'average', the particular event happens at a particular time or incurs a particular cost. The mean, median and mode of a dataset are collectively known as 'measures of central tendency', because these three measures focus on where the data are centred or clustered; that is, they are intended to represent the most typical or representative scores in a distribution.

MEASURES OF CENTRAL TENDENCY — MEAN, MEDIAN AND MODE

In the examples above, the *mean* is used to provide an indication of the average of all values. The mean is the sum of all values, divided by the number of observations or sample size. However, which measure of central tendency should we use if the data are not spread evenly; for example, if there are lots of high values or low values, or if the data are non-continuous (i.e. categorical)? In these instances, we need an alternative to the mean, to determine where the data are clustered or centred. This is where the 'median' and 'mode' are useful.

The *median* is the single value in a dataset that divides all of the values into half; that is, half the values will be above the median and half will be below the median. If we organise data relating to house prices into an ordered array, as shown in Table 9.3, then the median (shown in bold) is the rent at the midway point of the array.

Table 9.3 Sample data on cost of renting in a shared 3-bedroom house, with data arranged in order (odd number of houses)

House and location	Cost per week
House 2 - Bulleen	$95.00
House 5 - Shepparton	$95.00
House 10 - Heidelberg	$100.00
House 4 - Bendigo	$100.00
House 9 - Eaglehawk	$100.00
House 1 - Bundoora	**$105.00**
House 7 - Thornbury	$107.00
House 11 - Kingsbury	$115.00
House 3 - Ivanhoe	$145.00
House 6 - Fairfield	$160.00
House 8 - Eaglemont	$175.00

In the example in Table 9.3, there is an uneven number of rental houses (11) in the dataset. However, if we added another house in the data, to give a total of 12 houses, then the median would fall between the two scores in the middle of the range. In this case, the median would fall between the sixth and seventh numbers.

In this instance, the median is determined from the following formula:

(n + 1)/2

where 'n' represents the new sample size, in this case, n= 12, so the equation is:

(12+1)/2

Therefore, if there were 12 houses in our dataset, as shown in Table 9.4, the median now becomes $106 per week (i.e. midway between $105 and $107).

Table 9.4 Sample data on cost of renting in a shared 3-bedroom house, with data arranged in order (even number of houses)

House and location	Cost per week
House 2 - Bulleen	$95.00
House 5 - Shepparton	$95.00
House 10 - Heidelberg	$100.00
House 4 - Bendigo	$100.00
House 9 - Eaglehawk	$100.00
House 1 - Bundoora	**$105.00**
House 7 - Thornbury	$107.00
House 11 - Kingsbury	$115.00
House 12 - Macleod	$120.00
House 3 - Ivanhoe	$145.00
House 6 - Fairfield	$160.00
House 8 - Eaglemont	$175.00

The median can be used when ordinal, interval or ratio-scaled data are used, but it cannot be used for nominal data (i.e. categories) – for example, you cannot calculate the median from a dataset coded as male or female. Importantly, the median is selected as the measure of central tendency when ordinal, interval or ratio data are skewed (see description of skewed data in figure 9.2).

The median is often a more appropriate statistic to report than the mean when there is potential for outliers in the data set to impact the mean. For example, if you were interested in finding out what the average auction sale

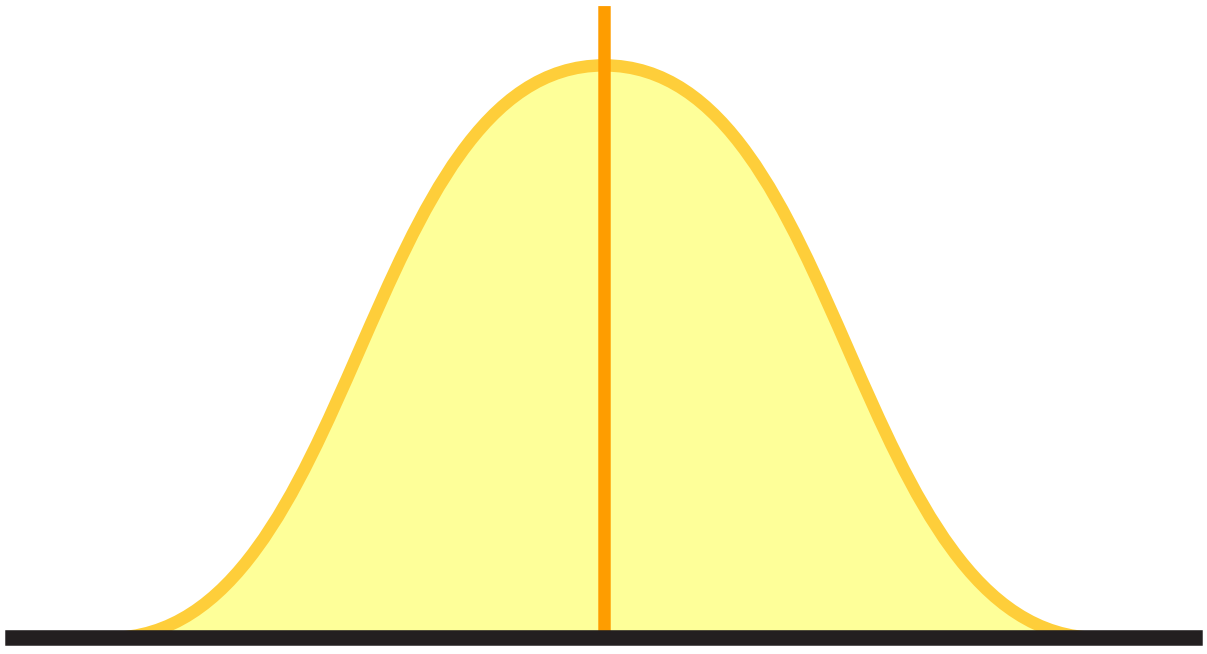
price was for houses sold in a particular city on any given Saturday and it just so happened that, while most sales fell within $500,000 to $3,000,000, there were three properties that sold for in excess of $15,000,000 then the mean sale price is going to be significantly affected by those sales. Using the median instead though, alleviates the considerable impact that these three outlier sales had on the mean.

The *mode* is used for nominal data – it represents the most frequently occurring score in a distribution. Technically, the mode could also be used for the data in Table 9.4 on house rental prices, where the most frequently occurring rental price is $100 per week. The mode can be used for any level of scaling (nominal, ordinal, interval or ratio). However, since it only takes into consideration the most frequently occurring number, the mode is generally not a satisfactory indication of central tendency.

MEAN, MEDIAN, MODE AND THE FREQUENCY POLYGON

When data have a normal distribution (i.e. are characterised by the bell-shaped curve shown in Figure 9.1), then the central blue line represents the value of the *median* and the *mode* and the *mean*.

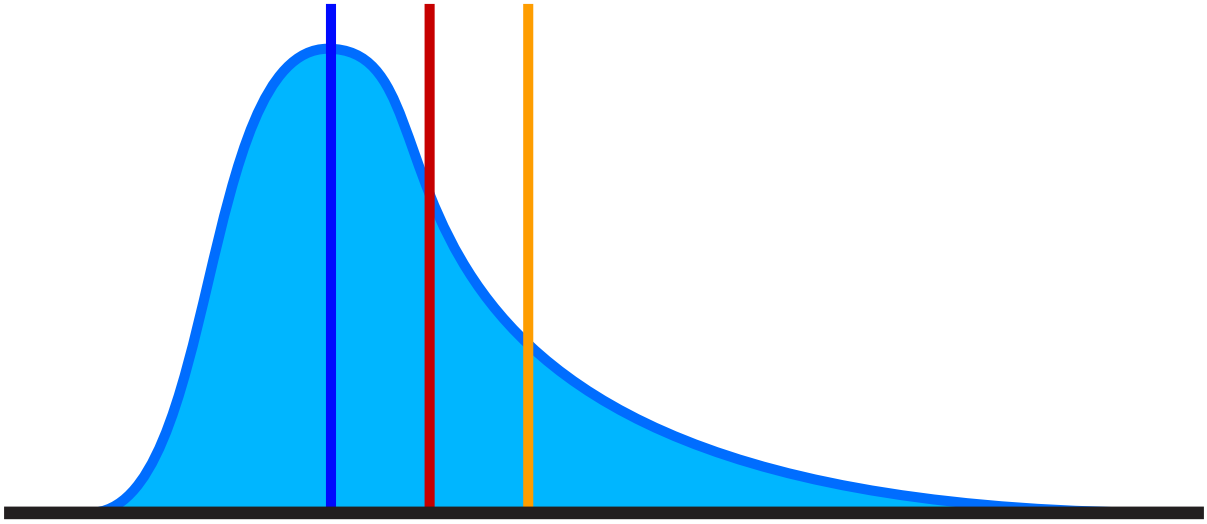
Figure 9.1 Data with normal distribution



Symmetrical
(Bell-shaped curve)

When data are positively skewed (i.e. are characterised by the long tail in the direction of the high scores shown in Figure 9.2), then the central blue line represents the value of the *mode*, moving to the right in red the *median* and then in green the *mean*.

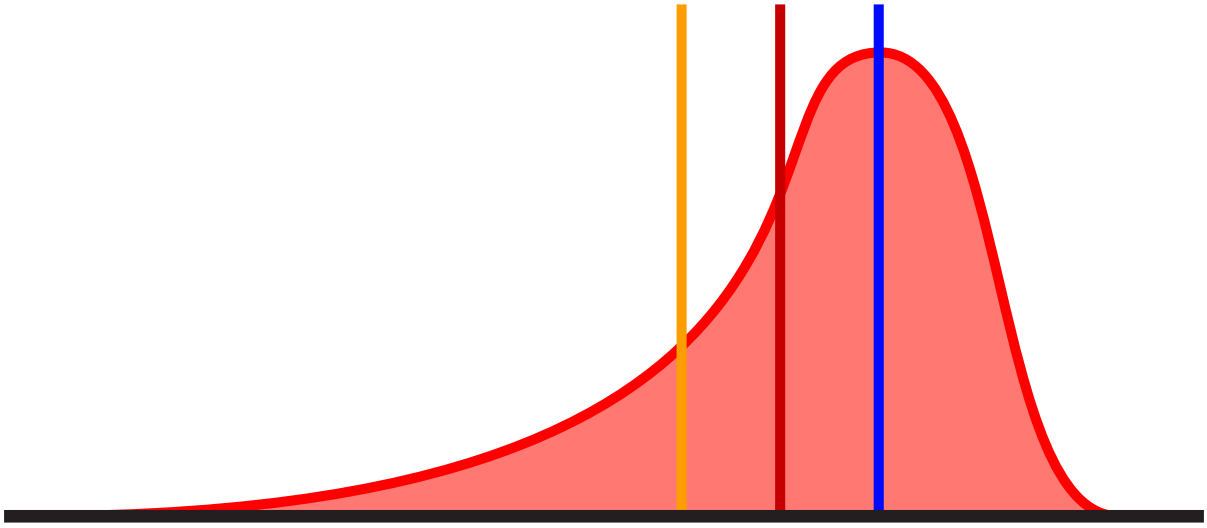
Figure 9.2 Data positively skewed



Mode < Median < Mean

When data are negatively skewed (i.e. are characterised by the long tail in the direction of the low scores shown in Figure 9.3), then the blue line represents the value of the *mode*, moving to the left in red the *median* and then in green the *mean*.

Figure 9.3 Data negatively skewed



Mean < Median < Mode

DESCRIPTIVE STATISTICS

The following presentation by Professor Geoff Cumming provides an overview of descriptive statistics:

https://doi.org/10.26181/5c118fc984634

Measures of variability – range, variance, standard deviation and interquartile range

If you were searching for a rental house (and had limited funds), you would want to know more about the market value of houses in a particular suburb than just the mean, median or mode. For example, you would probably like to know what the cheapest and most expensive options were, as well as how much the values were dispersed or spread out; that is you would be interested in the *variability* in the dataset.

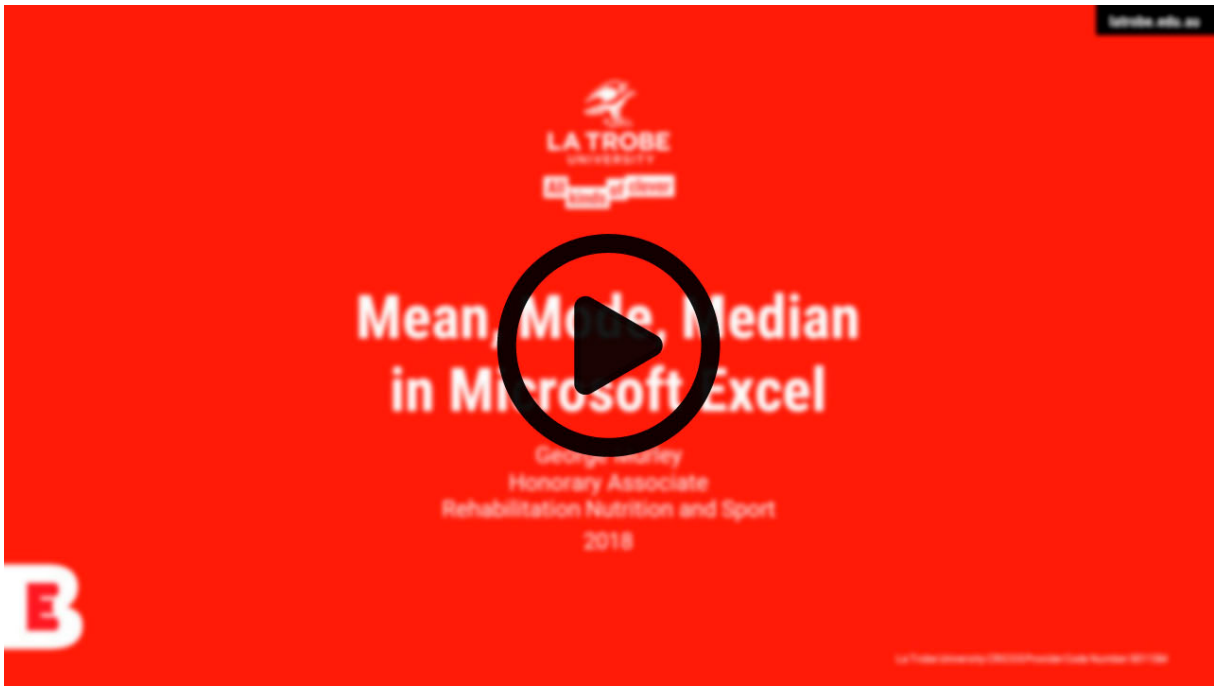
There are several methods for calculating variability. Perhaps the most basic is the *range*; that is, the lowest and highest values. For the rental houses in Table 9.4, the range is $95–$175. Using the range as an indicator of variability can be problematic when there are extremely low or high values – the 'outliers' in the dataset. For example, if we added a suburb such as Brighton into Table 9.4, then our range would be artificially large because Brighton (which is highly sought after) is not representative of most of the suburbs in the table. To overcome this problem, we often use other measures of variability such as the variance, the standard deviation and the interquartile range.

Computers and statistical software are used in calculating the variance, the standard deviation and the interquartile range, so you do not need to learn the formulae for these measurements of variability. The videos from Dr Murley and Professor Cumming explain how these measurements are computed.

The *variance* describes how far the numbers deviate from the mean (i.e. the average variability about the mean). For example, in Table 9.2, the rental house in Bundoora deviates +$13 from the mean of the dataset, and the rental house in Eaglehawk deviates –$13 from the mean of the dataset. To obtain the variance, you calculate the sum of the squared deviations about the mean and divide this by the number of cases – the short video clip below explains how this is calculated in Microsoft Excel. The higher the value of the variance, the greater the overall deviation from the mean (i.e. the greater the variability in the data). As highlighted in the video, the variance can 'overstate' the true spread of scores. Therefore, a more commonly used measure of variability for continuous data is the standard deviation.

The *standard deviation* is the square root of the variance – again, the video explains how this is calculated in Microsoft Excel. As with the variance, the higher the value of the standard deviation (relative to the mean), the greater the spread of data in a frequency distribution or frequency polygon.

In the following video Dr George Murley explains how to use Microsoft Excel to calculate the mean, median, mode and variance:



https://doi.org/10.26181/5c118f65f3d30

In the following video Professor Geoff Cumming presents an overview of standard errors:

https://doi.org/10.26181/5c118faa04547

How does standard deviation relate to normal distribution?

In the following video Professor Geoff Cumming illustrates the relationship between the standard deviation and the normal distribution:

https://doi.org/10.26181/5c118fbe860b2

Figure 9.4 shows the relationship between the standard deviation and the normal distribution. Can you see how the following observations are made from a normal (bell-shaped) distribution?

- the central two standard deviations (with one standard deviation either side of the mean) comprise 68% of the data
- the central four standard deviations (two standard deviations either side of the mean) comprise 95% of the data
- the central six standard deviations (three standard deviations either side of the mean) comprise 99% of the data

Figure 9.4 Relationship between the standard deviation and the normal distribution

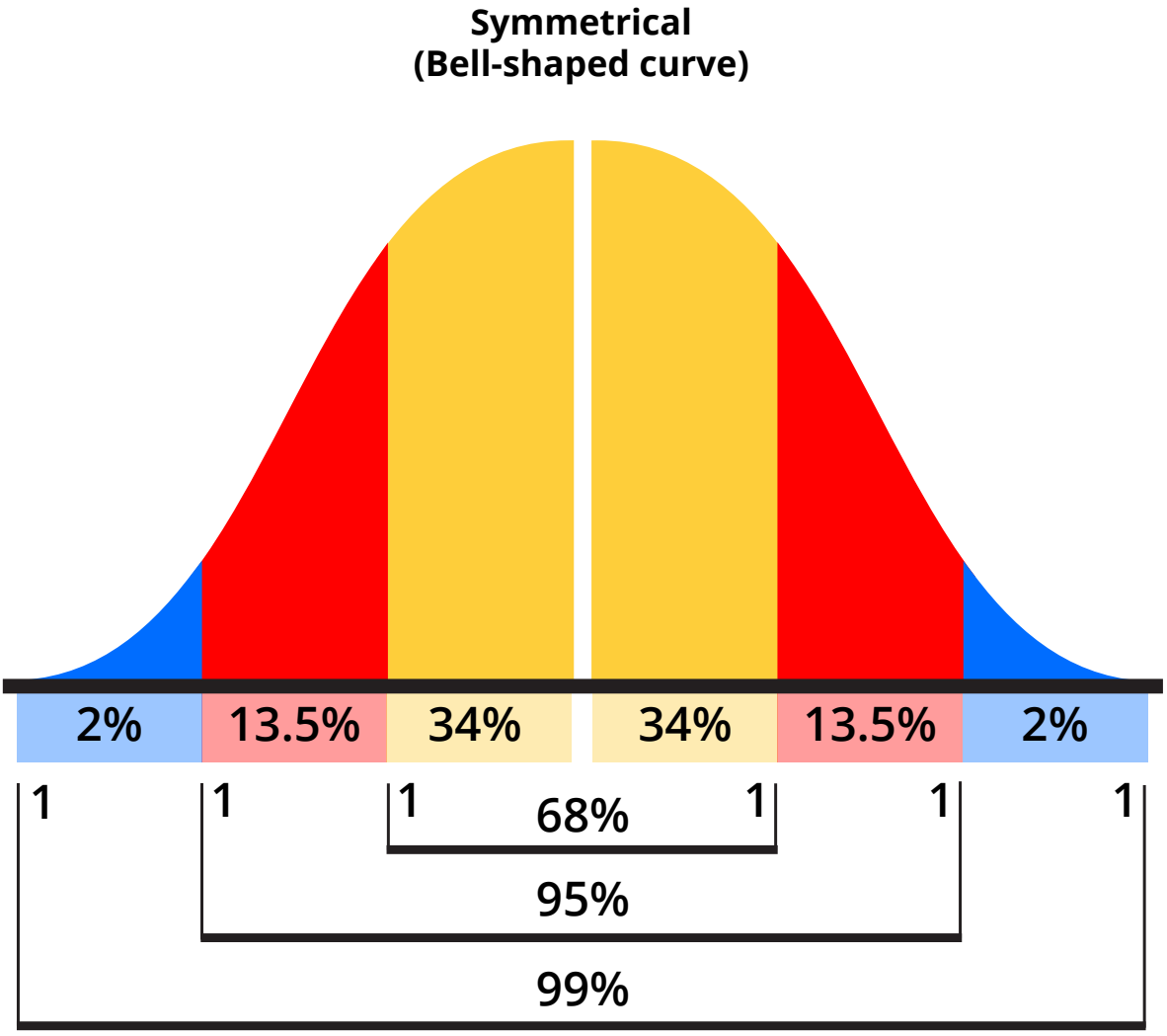


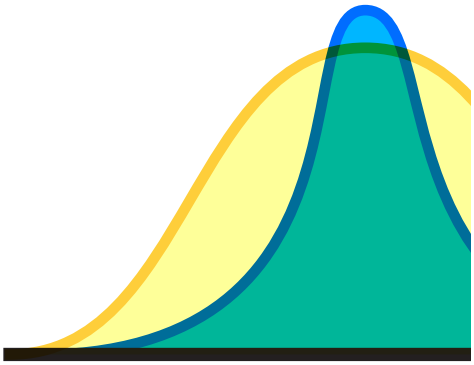
Figure 9.5 shows three frequency polygons. You can see that the variance for *minutes on the dance floor* is greater for the group that *consumed alcohol* (Polygon graph A) and for the group *not in a relationship* (Polygon graph B). You can also see that there is equal variance for the group comparison of caffeine (Polygon graph C). Essentially, the wider red polygon tracing indicates greater variance.

Figure 9.5 Three frequency polygons

Frequency Polygon A - A comparison of minutes spent on dance floor following alcohol consupmtion

Group 1 - No alcohol Consumed

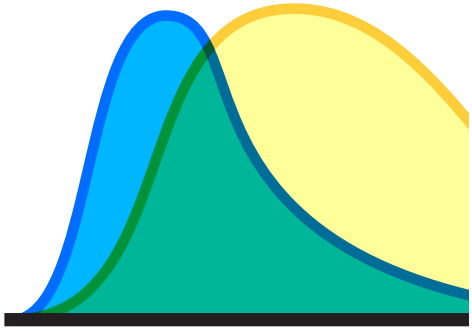
Group 2 - Alcohol Consumed



Frequency Polygon B - A comparison of minutes spent on dance floor depending on relationship status

Group 1 - In a relationship

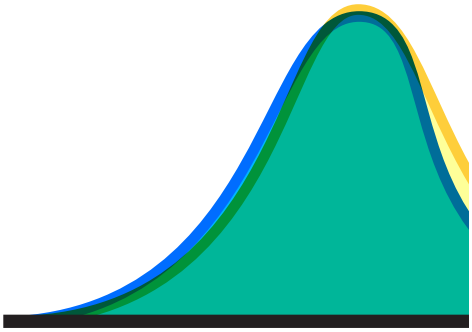
Group 2 - Not in a relationship



Frequency Polygon C - A comparison of minutes spent on dance floor following caffeine consupmtion

Group 1 - No caffeine consumed

Group 2 - Caffeine consumed



A – Effect of alcohol on minutes spent on the dance floor; B – effect of *relationship status* on minutes spent on the dance floor; and C – effect of *caffeine consumption* on minutes spent on the dance floor.

The *interquartile* and *semi-interquartile* ranges are used as the measures of dispersion (or variance) when the *median* has been selected as the measure of central tendency (i.e. usually when the data are skewed). The interquartile and semi-interquartile ranges are found by breaking the data into quarters (i.e. four equal parts, each containing 25% of the data), as follows:

- quartile 1 (Q1) is the lowest 25% of numbers
- quartile 2 (Q2) is the next 25% of numbers (up to the median)
- quartile 3 (Q3) is the second highest 25% of numbers (above the median)
- quartile 4 (Q4) is the highest 25% of numbers

Figure 9.6 shows how quartiles are distributed in normal and positively skewed data.

Figure 9.6 Frequency polygons illustrating how quartiles are distributed in normal (top) and positively skewed (bottom) data

What relevance do central tendency and variability have in health practice and research?

There are several reasons why we are interested in central tendency and variability in health practice and research. Central tendency is important because it allows us to understand what normal, common or expected, when we observe human characteristics. Understanding variability helps us to determine when an observation deviates from normal and by how much, and when this may become 'abnormal' or 'pathological'. Let us consider the serious public health issue of obesity as an example. The discussion below is related to body mass index (BMI) in the context of a sample population from the United States.

BODY MASS INDEX

BMI is used to estimate human body fat based on an individual's weight and height. It is calculated by dividing a person's mass (kg) by the square of the person's height (m):

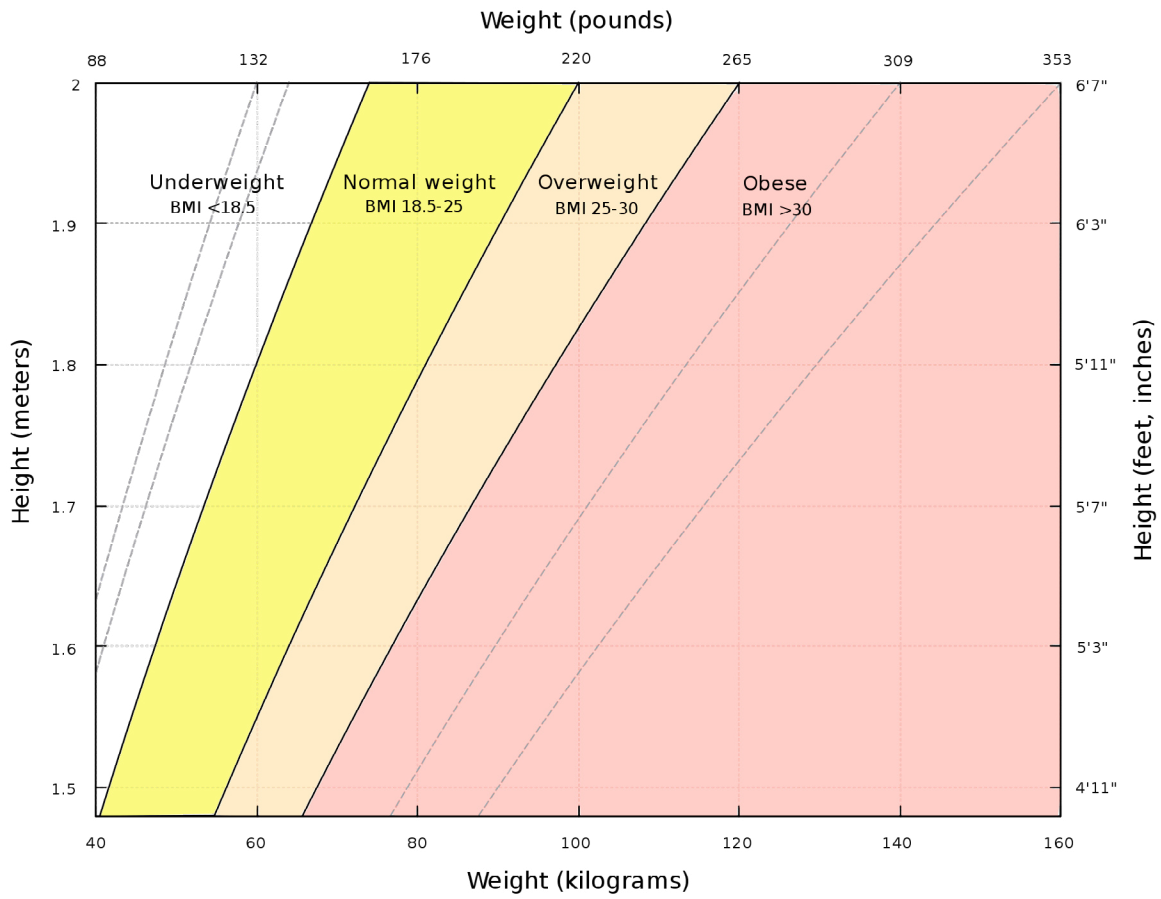
BMI = mass / (height × height)

If your mass was 70 kg and your height 1.77 cm then you could calculate your BMI as follows:

BMI = 70 / (1.77 x 1.77) = 70 / (3.1329) = 22.3435

The BMI value 22.3435 can be located on a BMI chart (Figure 9.7) to indicate where this value lies relative to the population average. You can see this value falls into the yellow band as 'Normal'.

Figure 9.7 BMI Classification



'BMI chart' from Wikimedia Commons used under CC BY-SA 4.0

OBESITY AND THE BMI DISTRIBUTION CURVE

Figures 9.8–9.10 are taken from a study by Penman and Johnson (2006), which investigated the change in the BMI distribution curve from 1990 to 2003, using cross-sectional data from an adult population in Mississippi.

View both the 1990 and 2003 distribution curves (and related data in the dialogue box) and see whether you can identify the following characteristics:

The mean, standard deviation and sample size for the 1990 dataset are 25.444, 4.8826 and 1498, respectively.

The mean, standard deviation and sample size for the 2003 dataset are 27.7308, 6.11952 and 4212, respectively.

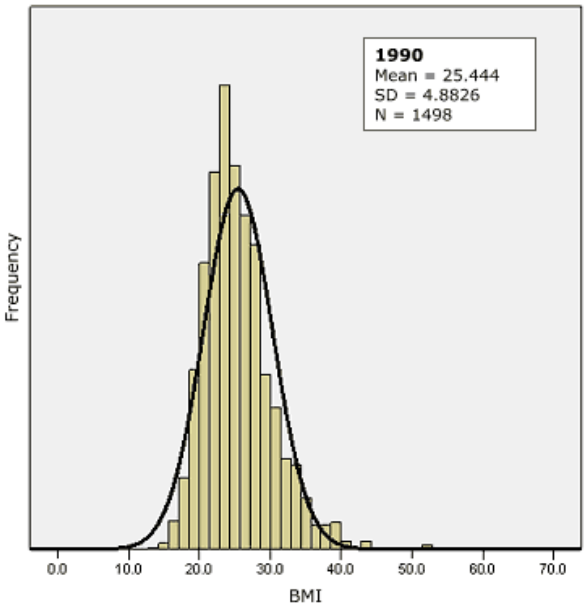
Generally, both distribution curves have the appearance of a normal-bell shaped curve (although technically they are both positively skewed).

The dataset from 2003 is slightly more positively skewed than the 1990 dataset.

The third graph is a line drawing of two overlapping distribution curves, which shows what happened from 1990 to 2003 – the spread became increasingly positively skewed over time (in plain English, the population became fatter!). Not only did the mean BMI increase (from 25.44 to 27.73), but there were more people with BMI values in the high tail of the curve; thus, in the 2003 curve, there were more people with BMIs greater than 40 than there were in 1990.

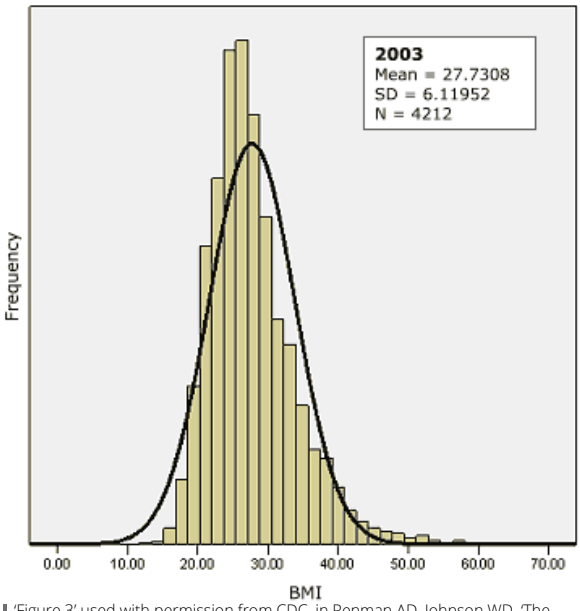
These results have implications for public health policy to reduce the prevalence of adult obesity, which is clearly worsening over time.

Figure 9.8 Body mass index distribution curve



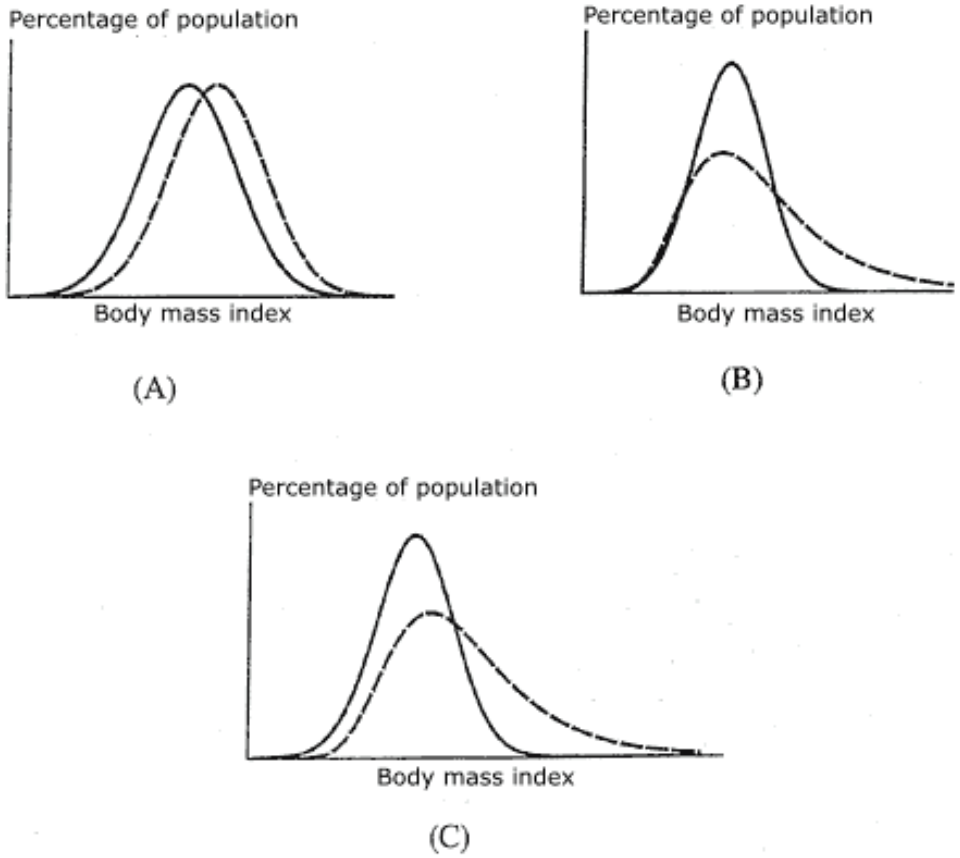
'Figure 2' used with permission from CDC, in Penman AD, Johnson WD. 'The changing shape of the body mass index distribution curve in the population: implications for public health policy to reduce the prevalence of adult obesity'. Available from: URL: http://www.cdc.gov/pcd/issues/2006/jul/05_0232.htm

Figure 9.9 Body mass index distribution curve



'Figure 3' used with permission from CDC, in Penman AD, Johnson WD. 'The changing shape of the body mass index distribution curve in the population: implications for public health policy to reduce the prevalence of adult obesity'. Available from: URL: http://www.cdc.gov/pcd/issues/2006/jul/05_0232.htm

Figure 9.10 Body mass index distribution curve



'Figure 1' used with permission from CDC, in Penman AD, Johnson WD. 'The changing shape of the body mass index distribution curve in the population: implications for public health policy to reduce the prevalence of adult obesity'. Available from: URL: http://www.cdc.gov/pcd/issues/2006/jul/05_0232.htm

Sampling

In the following video Professor Geoff Cumming discusses sampling:

https://doi.org/10.26181/5c118fb46dff4

Statistical inference

In the following video Professor Geoff Cumming discusses estimation:

https://doi.org/10.26181/5c118fa1e563f

Further reading

Descriptive statistics (2013) p.113. In. S. Polgar & S.A. Thomas, Introduction to research in the health sciences, 6th Ed., Churchill Livingstone Elsevier.

McKenzie, S. (2014). Vital statistics An introduction to health science statistics. Sydney: Elsevier Health Sciences APAC.

Salkind, N. J. (2016). Statistics for people who (think they) hate statistics. Sage Publications.

Chapter 10

Treatment effects: mean difference, odds ratio, risk ratio and confidence interval

Introduction and learning outcomes

Chapters 8 and 9 introduced scales of measurement and how various measurements are summarised through descriptive statistics. Descriptive statistics alone can provide useful information, but what is the next step if you want to compare the effectiveness of two therapies for the treatment of a condition? For example, if you suffer from anterior knee pain (i.e. pain at the front of your knee) when you run, how would you statistically compare the effectiveness of a muscle strengthening program (i.e. squats and balance exercises) to foot orthoses (i.e. shoe inserts or insoles that support the arch of the foot)? If one therapy was more effective than another, how would you determine whether that difference was large or small? To answer these questions, you would need to use a statistic that helps you to understand the size of the effect and whether the treatment or effect is worthwhile. This chapter will help you to interpret commonly used statistics related to treatment effects.

KEY LEARNING OUTCOME

Interpret from a journal article, statistics relating to treatment effects including mean difference, risk and odds ratios, and the confidence interval.

ENABLING OUTCOMES

Once you have completed this chapter, you should be able to:

- define treatment effects and identify common measures of treatment effect
- link a study design with the use of a 'risk measure' of treatment effect
- explain the meaning of the mean difference, standardised mean difference (SMD), relative risk and odds ratio in the context of the results of a research article
- explain the meaning of the CI associated with the treatment effect, in the context of the results of a research article

Mean difference and effect size: commonly used point estimates of treatment effect for continuous-scaled measurements

The term ‘estimate’ indicates that the data are obtained from a representative sample of the population. The ‘point estimate’ provides an estimate about what might be observed in the whole population from which the sample is derived.

MEAN DIFFERENCE

Perhaps the most basic and commonly reported measure of treatment effect size is the *mean difference*. This is the absolute difference between two sets of values. Table 10.1 presents the results of a trial involving the application of cream for the treatment of pain under the heel of the foot. The outcome measure is how many walking steps the participant can take before they begin to experience pain, after one week of daily administration of the cream.

Table 10.1 Results of a trial involving the application of cream for the treatment of pain under the heel of the foot

Participant	Baseline number of steps before treatment	Post treatment number of steps	Absolute difference between baseline and post treatment
1	32	32	0
2	34	37	3
3	51	53	2
4	12	15	3
5	27	29	2
6	26	29	3
7	11	13	2
8	34	39	5
9	43	50	7
10	15	18	3
Mean	**28.5 (Mean 1)**	**31.5 (Mean 2)**	**3**
		Absolute mean difference	**3 steps**

You can see from Table 10.1 that, after one week of daily administration of the cream, participants were able to take (on average) three more steps before they experienced pain. The example of the mean difference above is easy enough to comprehend, but what does 'three extra steps' indicate in terms of the magnitude – is this a small, moderate or large change? To answer this question, we use one of the most practical and standardised measures of treatment effect for comparing two means: the SMD.

STANDARDISED MEAN DIFFERENCE

The SMD (also referred to as the effect size or Cohen's *d*) expresses the absolute change relative to the standard deviation. There are several ways to calculate the SMD, the most common of which involves taking the absolute difference (mean difference between the **experimental** condition and the control) and dividing this by the standard deviation (either the pooled standard deviation, or the standard deviation of the baseline scores):

$$\frac{(Mean\ 1 - Mean\ 2)}{pooled\ standard\ deviation}$$

$$\frac{(31.5 - 28.5)}{14} = 0.21$$

At this point, you may still be wondering why the value 0.21 is more useful than the information about the 'three extra steps' derived from the mean difference. Remember that we want to know something about the *relative size* of the effect (i.e. how 'big' a difference is taking 'three extra steps'). Well, because the SMD is standardised, the effect size is telling you how much change has occurred relative to the standard deviation. An SMD of:

- 0.20 or less represents a small change
- 0.50 represents a moderate change
- 0.80 represents a large change

Therefore, in the example above, an improvement of three extra steps is considered a small to moderate effect.

A useful feature of the SMD is that when an experiment has been replicated, it is easier to compare the different effect size estimates to those of other studies that used a similar method.

Some issues may prevent the use of the effect size calculation. Such issues include skewed data and situations where the means for the treatment conditions have appreciably different standard deviations (e.g. if the post-treatment score standard deviation is appreciably different from the baseline score).

Odds ratio and relative risk: commonly used point estimates of treatment effect for discontinuous-scaled measurements

Two popular point estimates of treatment effect are the odds ratio (OR) and relative risk (RR), which are used to compare the risk in two different groups of people. In health research, groups of people (e.g. smokers) are compared to other groups (i.e. non-smokers), to see whether belonging to a group increases or decreases a person's risk of developing certain diseases (e.g. lung cancer). OR and RR are usually interpreted as being equivalent, but there are some minor exceptions, as discussed later in this chapter.

Both OR and RR are referred to as *risk ratios*, and in both cases the values range from zero to infinity. Importantly, values greater than 1.0 indicate *increased* risk, whereas values less than 1.0 indicate *reduced* risk. Values equal to 1.0 indicate that the risk is no better than chance (therefore 'no effect' is detected in terms of the point estimate of treatment effect).

Table 10.2 Value of odds ratio or relative risk ratio

Value of odds ratio or relative risk ratio	Definition
0 to <1	Reduced risk of event
>1	Increased risk of event
=1	No effct - risk of event is no better than chance
0	Event will never happen
Some examples:	
0.2	Reduced risk of event
1.01	Increased risk of event
7.0	Increased risk of event
0.9	Reduced risk of event

ODDS RATIO

The OR is a way of representing probability. This is especially familiar in betting; for example, the odds that flipping a coin will produce a 'heads' are 1 in 1. Therefore, the 'odds of an event' is the number of cases who experience

the event of interest, divided by the number of those who do not experience the event of interest. As stated above, the OR is expressed as a number from zero (i.e. the event will never happen) to infinity (i.e. the event is certain to happen). Odds are fairly easy to understand when they are greater than one, but are harder to understand when the value is less than one.

Putting this into context, imagine a trial investigating the effect of zinc (vitamin) supplementation on the incidence of the common cold. **You are not expected to remember the equation, this example is given purely to help you to understand OR.** Table 10.3 is a '2×2 table' that shows the number of events; that is, the number of individuals who developed a cold, and the number of individuals who did not develop a cold, including whether they belonged to the zinc or **placebo** group).

Table 10.3 Treatment and outcome for zinc and common cold

Treatment	Outcome	
	Number of individuals who developed a cold (positive event, meaning the event occured)	Number of individuals who DID NOT develop a cold (negative event, meaning the event did not occur)
Zinc group	50	125
Placebo group	100	75
Total	**150**	**200**

Thus, the odds that a person with a cold was taking zinc can be calculated as:

$$\frac{50/150}{100/150} = 0.50$$

And the odds that a person without a cold was taking zinc can be calculated as:

$$\frac{125/200}{75/200} = 1.67$$

Therefore, the OR that someone with a cold is taking zinc compared to a person without a cold taking zinc can be calculated as:

$$\frac{0.5}{1.67} = 0.30$$

The OR of 0.30 is less than 1, indicating reduced odds or reduced risk. It means that a person with a cold is 0.30 times as likely to have taken zinc, than a person without a cold. Expressed in that way the OR is quite difficult to comprehend. We can break it down further to state that from 13 people who presented with a cold, you would expect that 3 would be taking zinc and the other 10 would not.

Let us consider another hypothetical example, this time to see how an increase in the OR can occur. Table 10.4 is another '2x2 table' – this time summarising the results of a trial investigating injury rates in a group of people who ran a marathon in running shoes and another group who ran barefoot.

Table 10.4 Treatment and outcome for footwear and injury rates

Treatment	Outcome	
	Number of individuals who developed an injury (positive event, meaning the event occurred)	Number of individuals who DID NOT develop an injury (Negative event, meaning the event did not occur)
Footwear running group	155	165
Barefoot running group	75	240
Total	**230**	**375**

Thus, the odds that an injured person was running with footwear can be calculated as:

$$\frac{155/230}{75/230} = 2.07$$

And the odds that a non-injured person was running with footwear can be calculated as:

$$\frac{165/375}{240/375} = 0.67$$

Therefore, the odds of 'being injured' when running the marathon in footwear compared to a 'non-injured' person running with footwear can be calculated as:

$$\frac{2.07}{0.69} = 3.01$$

An OR of 3.01 is greater than 1, indicating increased odds or increased risk. It means that a person who is injured is 3.01 times more likely to be running with footwear, compared with a 'non-injured' person (note: this is an imaginary experiment, not a real one!).

ORs are presented in a range of scenarios or study methods including case–control and prospective studies. They are not only used to estimate treatment effects but are also used to estimate the odds of developing a disease (e.g. heart disease, knee pain or diabetes) in the presence of a particular characteristic (e.g. smoking or not smoking, flat-feet or normal feet, or physical exercise or no physical exercise).

RELATIVE RISK

As you will now realise, several types of 'risk' are reported in the health research literature; for example, relative risk reduction, absolute risk reduction and number needed to treat. Here, we focus on RR, which like the OR, is used to compare the risk in two different groups of people. More precisely, the RR is the ratio of the incidence in people with the risk factor (exposed persons) to the incidence in people without the risk factor (non-exposed persons). Therefore, RR is not only used to estimate treatment effects (e.g. using some type of therapy), but also to estimate the risk for developing a disease in the presence of a particular characteristic.

Let us consider the example given earlier in Table 10.2 relating to taking zinc for the common cold. The risk that a person taking zinc had a cold can be calculated as:

$$\frac{50}{50 + 125} = 0.29$$

The risk that a person taking a **placebo** had a cold can be calculated as:

$$\frac{100}{100 + 75} = 0.57$$

Therefore, the 'relative' risk that a person taking zinc would develop a cold can be calculated as:

$$\frac{0.29}{0.57} = 0.5$$

From the calculation above you can see that a person taking zinc, is half as likely to develop a cold as another person taking the **placebo**, which you

probably expected from reading the example provided in the context of an odds ratio. You might also notice that the value of the RR (0.5) is 'higher' than that of the OR (0.3). This means that the OR has overestimated the size of the effect (a lower value is a larger effect). For this and other reasons (**too lengthy to be** discussed here), there is a 'push' for researchers to use RR rather than OR for reporting treatment effects. Despite this, you will come across both RRs and ORs when reviewing evidence of treatment effects from health research literature.

Take a deep breath! You now know the main statistics used to report point estimates of treatment effects (mean difference and SMD for continuously scaled outcome measurements, and risk ratios – either RRs or ORs – for non-continuously scaled outcome measurements). These are 'estimates' because they are calculated from a sample of people who represent the population. So the true treatment effect might differ in the whole population of interest. If you repeated the experiment (either with the whole population or another sample) you can expect the effect size to be slightly different. But how different – 10%, 20%, 30% or 40% different? This leads us to the CI.

Confidence intervals

When reviewing evidence from a research article, you need to make some decisions about the *precision* of the point estimate of treatment effect. As stated above, if an experiment is repeated, the point estimate may be smaller or larger than the original study. As a health practitioner, you should be asking, 'what might be the smallest treatment effect my client is likely to experience if they receive a particular therapy?' The answer to this question is found in the CI. The CI is the key to interpreting treatment effects, and is paramount when deciding whether the treatment effect makes a treatment worth implementing in clinical practice.

The CI is a range, either side of the point estimate that tells you how much the point estimate may vary in the population. It is sometimes described as a *margin of error*. Confidence 'limits' are simply the extreme ends of the CI – the highest and lowest values of the interval.

To calculate the CI you need to know three things: the sample size, the standard deviation and the 'level of confidence'. The level of confidence part relates to probability; that is, the probability that the point estimate is contained within the interval. Commonly reported CIs are 95%, 98% and 99%, where, for example, a 95% CI indicates that there is a 95% probability that the point estimate is contained within the 95% CI.

As with all of the statistics presented here, you are not expected to remember the equation or how to perform the calculations. Instead, the aim is for you to be able to understand and interpret these statistics from research articles.

In the following video Professor Geoff Cumming presents an overview of CIs:

https://doi.org/10.26181/5c118f99a933e

The rest of this section provides some hypothetical examples. We need to be specific about what type of point estimate of effect is being used; therefore, the examples are separated into those relating to continuous measures of effect and those relating to discontinuous measures of effect.

CONFIDENCE INTERVALS FOR THE MEAN DIFFERENCE AND STANDARDISED MEAN DIFFERENCE

Imagine that a research article reports the findings from a trial investigating the effectiveness of caffeine for overall performance on university exam results, and has the following findings:

> students who consumed caffeine scored 15 out of 100 points *higher* on the test than students who consumed a decaffeinated **placebo**; therefore, the mean difference was 15 points
>
> let us say that the standard deviation for the mean difference was 10, that the sample size was 50 students and that we have calculated the 95% CI as ± 2.77

If we subtract 2.77 from the mean difference we get the lower limit of the CI (12.23), and if we add 2.77 to the mean difference we get the upper limit of the CI (17.77); therefore, our 95% CI relative to the mean difference is 12.23–17.77.

This CI indicates that we can be 95% certain that the true performance (i.e. the point estimate of the treatment effect) in the population is somewhere between 12.23 and 17.77. That all sounds good – our present experiment is indicating that supplementing with caffeine gives a 95% chance that, on average, you could expect an improvement in your grade of at least 12.23 points, and possibly as high as 17.77 points!

If we calculate the SMD for the data above we get a value of 1.5, which is a 'large effect size'. The 95% CI for the SMD is 1.19–1.81. Again, we can say that supplementing with caffeine gives a 95% chance that, on average, you could expect a large effect (or better) in terms of improvement in performance.

It would be nice if all research articles you read reported such compelling findings. The reality, however, is that the CI often tells us a very different story. One factor that has a significant effect on the CI is the sample size. Let us manipulate the data from the caffeine exam trial above to explore the effect of sample size on the width of the CI; we will do this by:

- reducing the sample size from 50 to 10 students
- increasing the 95% CI from ± 2.77 (with n=50) to ± 7.15 (n=10)

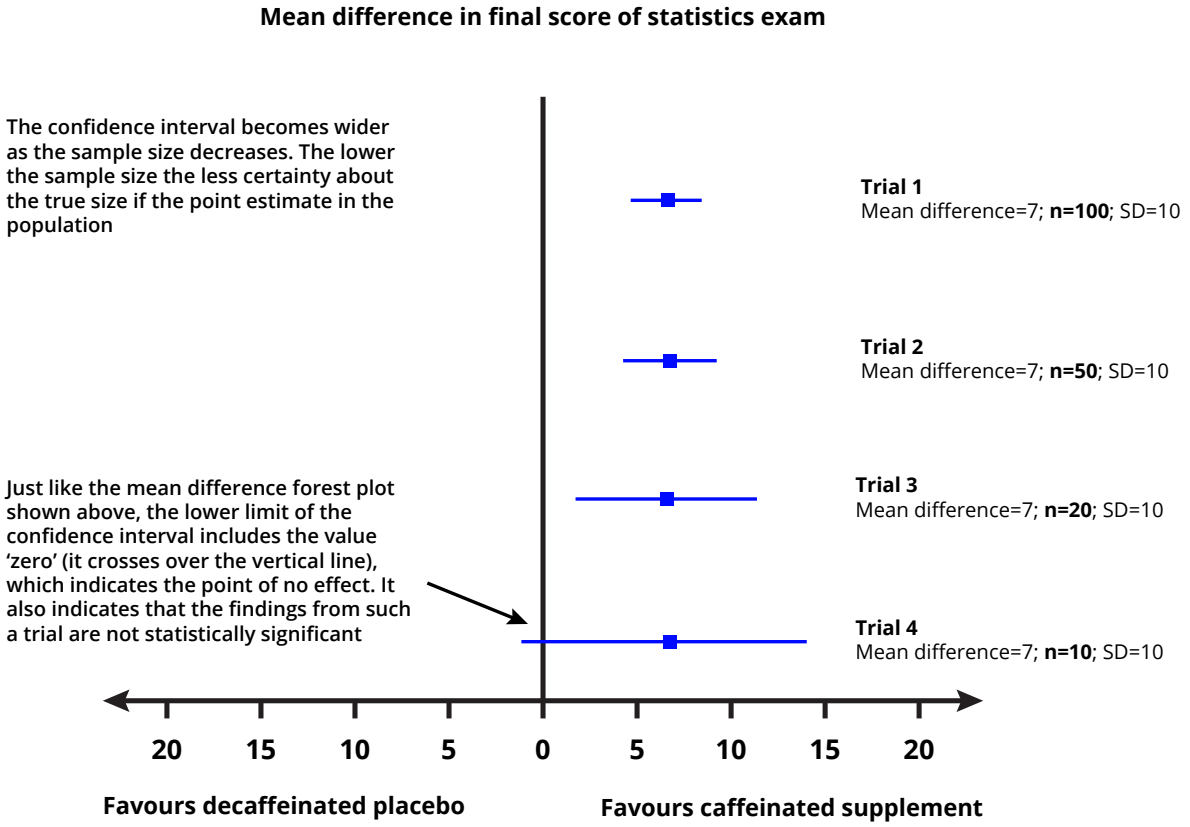
If we subtract 7.15 from the mean difference we get the lower limit of the CI (7.85), and if we add 7.15 to the mean difference we get the upper limit of the CI (23.15); therefore, our 95% CI relative to the mean difference is now 7.85–23.15.

This CI indicates that we can be 95% certain that the true performance (i.e. the point estimate of the treatment effect) in the population is somewhere between 7.85 and 23.15. Can you see what has happened? By reducing the sample size, the 95% CI has become much wider. We are now less certain about the precision of the point estimate of treatment effect (15 points). The SMD is still 1.5, which is still a large effect size; however, the 95% CI for the SMD is now 0.51–2.49, which is a moderate to large effect.

Figure 10.1 is a hypothetical forest plot with a series of point estimates and their 95% CIs. All of the point estimates have a value of 7 points with a standard deviation of 10, but the sample size varies for each point estimate. There are two key concepts for you to take away from reviewing this forest plot. First, the CI becomes wider as the sample size decreases; that is, the lower the sample size, the less certainty about the true size of the point estimate. Second, you can see from the CI with the smallest sample size (n=10)

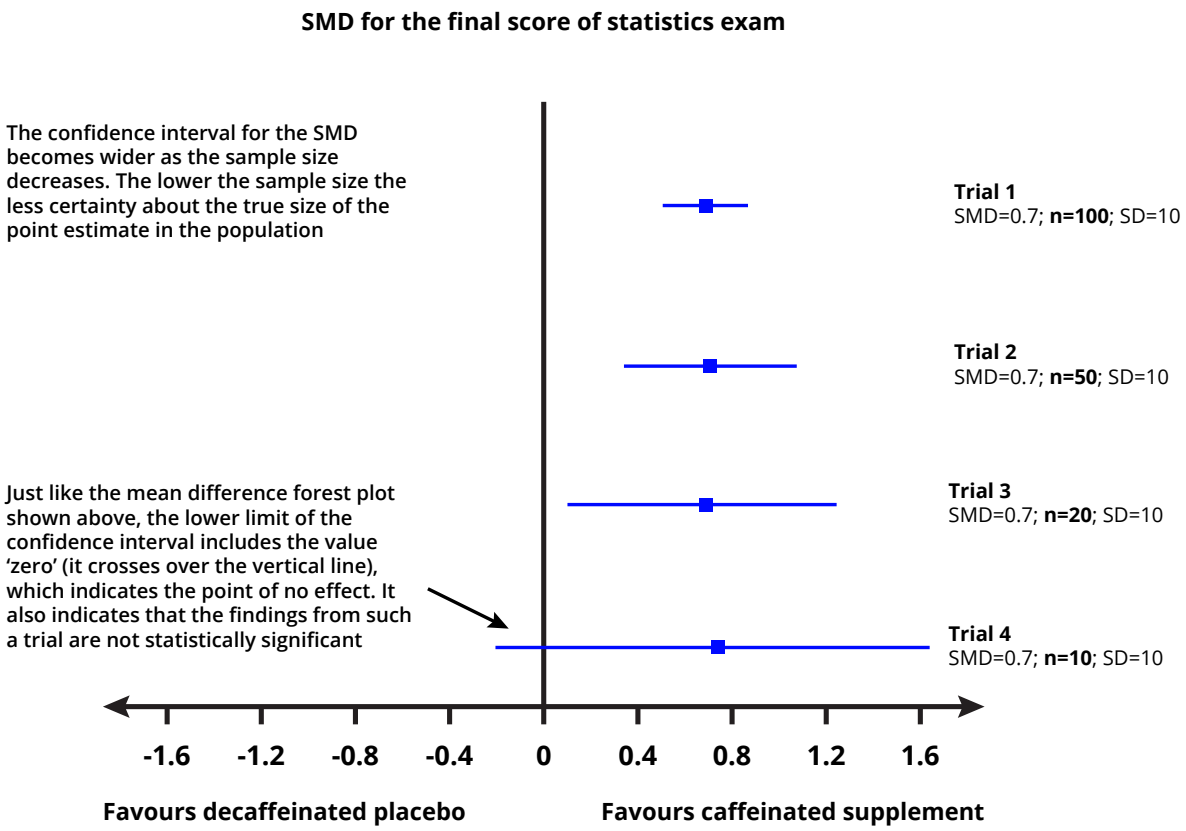
that the 95% CI actually crosses over the vertical line; that is, the lower limit of the CI includes the value zero (0). This indicates that there is a 95% chance that the true estimate of effect is located somewhere between favouring the decaffeinated **placebo** and favouring the caffeinated supplement.

Figure 10.1 Mean difference in final score of statistics exam



Now let us convert the values in the forest plot above into SMDs, and plot the CIs on another forest plot (Figure 10.2). Note that the unit of measurement will become SMD rather than mean difference.

Figure 10.2 SMD for the final score of statistics exam



The examples above were calculated using 95% CIs. If we calculated 98% or 99% CIs for the same data, this exercise alone would make the CIs wider. The choice of using either a 95%, 98% or 99% CI comes down to how certain you need to be about the potential range of the point estimate in the population.

In summary, the larger the standard deviation (or variance), the wider the CI, and therefore the less certainty there is about the true estimate in the population.

CONFIDENCE INTERVALS FOR RISK RATIOS: ODDS RATIO AND RELATIVE RISK

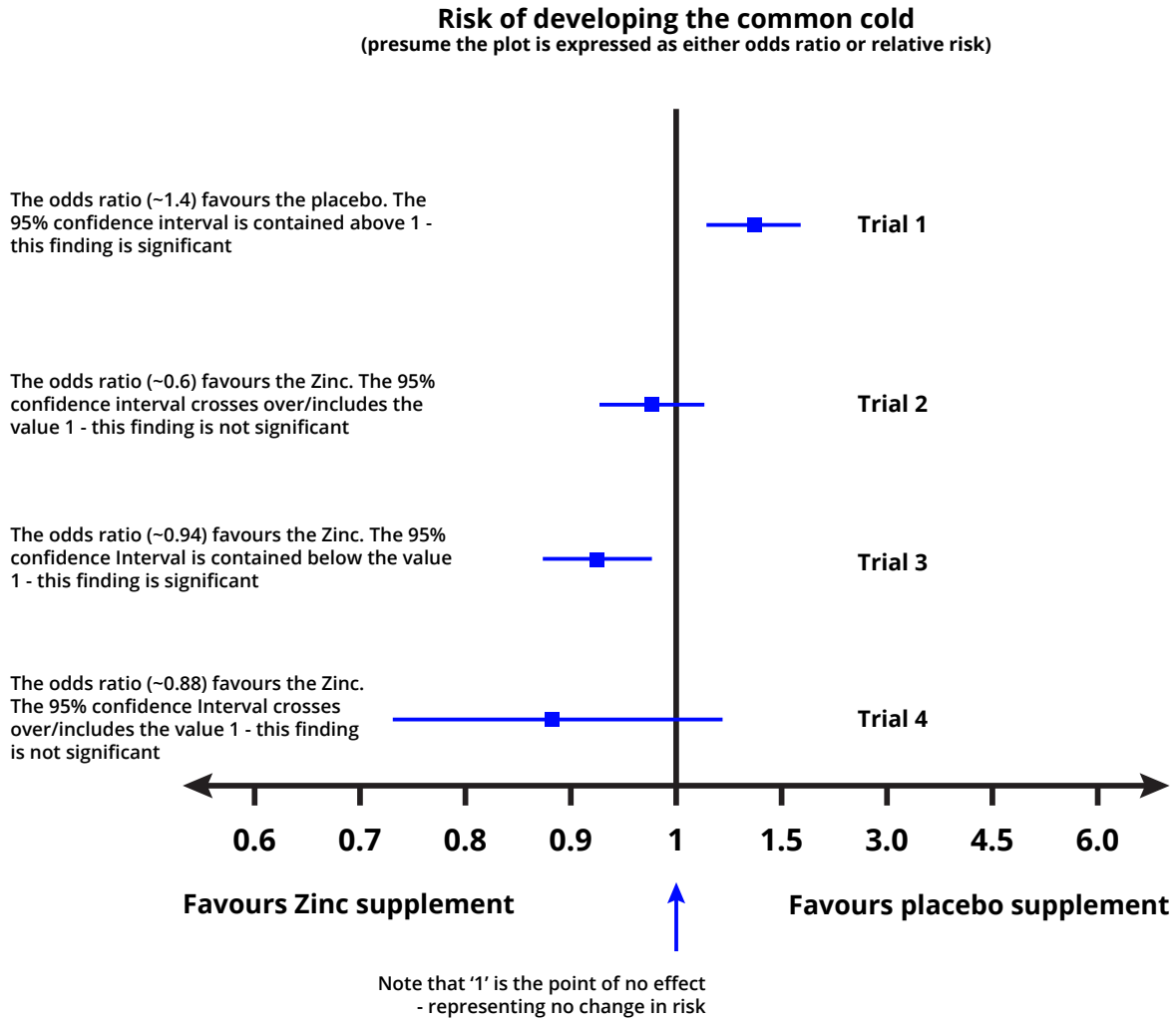
Nearly all of the principles for the CI remain the same when interpreting CI in the context of risk ratios, except for one important difference. In the first part of this chapter we noted that risk ratios with a value of 1.0 indicate a point of no effect (i.e. no change in risk). Of course, this also applies to the CI for the risk ratio (i.e. to odds ratio and relative risk).

First, let us consider the zinc study results presented figure 10.3 below. The odds ratio that someone with a cold was taking zinc compared to the **placebo**

was 0.30. The 95% CI for this point estimate is 0.18–0.53. Thus, there is a 95% chance that the true odds ratio in the population lies somewhere between 0.18 and 0.53.

This finding is significant because, at worst, you would expect the odds ratio to be 0.53, which is less than the point of no effect for measures of risk (1.0). What if the 95% CI were 0.18–1.1, would this be a significant finding? The answer is no, because the 95% CI now crosses the value 1.0 (the critical value indicating the point of no effect or no change in risk). The concept of the 95% CI and the point of 'no effect' is further explained through the use of the hypothetical forest plot in Figure 10.3.

Figure 10.3 Risk of developing the common cold



SUMMARY OF CONFIDENCE INTERVALS

To summarise the key points about the CI:

- the CI is a range, either side of the point estimate, that tells you how much the point estimate may vary in the population
- to calculate the CI for the mean difference and SMD, you need to know three things: the sample size, the standard deviation and the 'level of confidence' (i.e. 95%, 98% or 99%)
- a 95% CI of '5–10' indicates that there is a 95% probability that the true population 'treatment effect' is contained within the values 5 and 10
- the smaller the sample size, the wider the CI, and the larger variance in the data, the wider the CI
- when the lower limit of the CI crosses includes the value 0 (for mean difference and SMD) it indicates that the findings from such a trial are not statistically significant
- when the lower or upper limit of the CI includes the value 1 (for risk ratios) it indicates that the findings from such a trial are not statistically significant

In the following video Dr Karl Landorf discusses treatment effects and CIs:

https://doi.org/10.26181/5c1190111444a

Further reading

Sedgwick, P. (2015). Confidence intervals, P values, and statistical significance. BMJ: British Medical Journal, 350(Feb27 1), H1113-h1113.

Probability and confidence intervals (2013). p. 149. In. S. Polgar & S.A. Thomas, Introduction to research in the health sciences, 6th Ed., Churchill Livingstone Elsevier.

Salkind, N. J. (2016). Statistics for people who (think they) hate statistics. Sage Publications.

Chapter 11

Probability, statistical significance and power

Introduction and learning outcomes

Chapter 10 introduced various measures of treatment effect and the important concept of the CI. By now you can probably appreciate that when research articles present a point estimate and the associated CIs, this is usually sufficient to make decisions about whether the findings are 'statistically significant' or whether two groups of data are different.

In your quest for evidence, you will read many research articles that use only a **p-value** to explain whether a statistical difference exists between therapies or observations. There are several limitations and even flaws with using p-values to interpret results from research articles. One such problem is that the selection of a critical p-value is often fairly arbitrary. For example, if a study selects a critical p-value of 0.05 and the experiment produces a finding with a p-value less than 0.05, this is interpreted as 'statistically significant', but if the p-value is 0.051, then the finding is no longer statistically significant. Clearly, using p-values to make a decision about statistical significance is problematic. Despite this, you need to understand what p-values mean, because most journal articles still rely heavily on p-values to indicate statistical significance.

This chapter explains how to interpret p-values from scientific articles. It is highly recommended that you read through this chapter at least twice and undertake the activities at the end, to ensure that you have a clear understanding of the content.

KEY LEARNING OUTCOME

Identify and explain statistical significance in the context of a research article.

ENABLING OUTCOMES

Once you have completed this chapter, you should be able to:

- define statistical significance, type I and type II errors, and power
- explain the purpose of **null hypothesis** significance testing
- identify values in a research article that indicate statistical significance
- explain why sample size calculations are important, and what **variables** are used to calculate sample size

Null hypothesis significance testing – probability and p-values

A **hypothesis** is an idea or explanation for something that is based on known facts but has not yet been proved. To prove or disprove the explanation, it is necessary to undertake an experiment using a **scientific method** (i.e. a research experiment). We call this process 'hypothesis testing'.

NULL HYPOTHESIS SIGNIFICANCE TESTING

Most of the research articles you review will contain some level of structured hypotheses – perhaps as part of the aims of the study. Sometimes the **hypothesis** is not clearly stated, but you can deduce what it might be from the statistical approach presented in the results.

There are two types of statistical hypotheses:

> The **null hypothesis** – is usually the hypothesis where results were obtained from the same sample and there is no real difference between groups or observations.
>
> The ***alternative hypothesis*** – is the hypothesis where sample observations are influenced by some non-random cause, and there is a real difference between the groups due to some systematic cause. The alternative hypothesis is therefore the counterpart of the null hypothesis.

'Null hypothesis significance testing' means that statistical testing using **p-value**s is used to decide whether to reject or retain the null hypothesis. The outcome is dichotomous – there is either a difference or there is not. You might think it strange that the research question is geared around whether or not the null hypothesis is accepted or rejected. This process is described further below.

In the following video Professor Geoff Cumming provides an overview of **null hypothesis** significance testing and statistical significance:

https://doi.org/10.26181/5c118f8ed2302

WHAT DOES THE P-VALUE INDICATE?

P-values provide additional information to help us determine whether results are statistically significant. Technically, p-values help us to decide whether or not to reject the null hypothesis. This might seem somewhat backwards thinking, in that we are assuming there is no difference and then hoping to demonstrate that a difference exists.

Definition: *The p-value first assumes that the null hypothesis is true, and then indicates the probability of obtaining the observed difference (or a larger difference). In simpler terms, the p-value is the probability that the observed result (or greater) occurred by chance alone.*

Let us imagine that a research article reports that Group A performed 12 points higher on a test than Group B, and that the p-value is 0.01 (i.e. p=0.01). This p-value indicates that there is a 1% chance that the observed difference (12 points or more) would be due to chance. In other words, it is unlikely (i.e. only a 1% chance) that the difference between Group A and B would be as large as 12 points due to chance alone.

The fact that we describe the probability (i.e. the p-value) relative to the null hypothesis is why statistical testing of this type is called 'null hypothesis significance testing'. It represents the strength of the evidence provided by the sample data in support of the null hypothesis. The p-value is a probability and is therefore indicative of how likely an event is to occur (e.g. in this case, the probability of obtaining 12 points by chance alone).

As the **p-value** approaches a value of 1, this adds support to the **null hypothesis**. However, as the p-value approaches zero, this adds support to the alternative **hypothesis**. You will notice that many scientific journal articles set the critical value to 0.05 (i.e. 5%), which is referred to as the critical level of significance or alpha (sometimes written as '*α*'). You will come across articles that set an alpha value of 0.10 or less. By definition, alpha is also the probability of performing a type I error, which is described below (this is why you need to read this topic twice!).

So, a p-value of less than 0.05 provides support for the alternative hypothesis (i.e. there is a difference between groups), and means that we reject the null hypothesis in favour of the alternative hypothesis. If the p-value is less than 0.05, we can say there is a statistically significant difference between the two conditions tested. Moreover, if we are talking about the results from a clinical trial, we could also say that Condition A is 'unlikely' to be less effective than Condition B.

In the following video Professor Geoff Cumming provides an overview of the p-value:

https://doi.org/10.26181/5c118f869d4aa

Can the conclusions from null hypothesis significance testing be wrong?

Since **hypothesis** tests are based on estimates of probability, their conclusions can be erroneous. In this regard, there are two types of error, type I and type II, as discussed below.

TYPE I ERRORS

When you are reviewing a research article, you need to remember that sometimes researchers can report significant findings that in fact have only occurred by chance. The more 'comparisons' a researcher makes, the more likely they are to detect a 'significant finding'. For example, if a researcher conducted a trial and compared hundreds of outcome measurements – such as pain, function and well-being scores – then there is a reasonable chance that at some point the groups in the trial will be different due to chance alone (i.e. a random finding rather than a systematic finding). This type of result is called a type I error, and it occurs when a true **null hypothesis** is rejected (i.e. when the authors conclude there is a 'difference' when in fact no difference exists). Even though the researchers detected a significant finding, in actual fact this occurred only by chance alone and there is no real difference. The problem here is that there is no way of knowing whether the result was by chance alone or was a true systematic difference; in this situation, you must rely on making a good judgement from the statistics.

TYPE II ERRORS

Sometimes, in contrast to a type I error, there are real systematic differences between groups that are not detected. This is a type II error, and it involves failing to reject the null hypothesis when it is false. Type II errors usually occur when there is a small sample size, which leads to wide CIs that cross the point of 'no effect'. In these situations, there is in fact a difference, but the sample size is insufficient to power the study.

To explain this situation further, sometimes it is concluded that a study has failed to show a difference as indicated by the **p-value** greater than 0.05 (i.e. $p>0.05$). However, there is an old saying that, *'absence of evidence is not evidence of absence'*. When we are told, 'there is no difference between A and B', we should first ask whether absence of evidence means simply that there is no information at all. There may in fact be an important difference between A and B, but this may have been missed because of a small sample size; that is, there may have been a type II error of missing a statistically significant difference when one exists.

Power and sample size estimations

As we have discussed, most studies involve a sample of participants rather than an entire population of interest – we study a sample of participants to draw inferences about the whole population. Power and sample size estimations are measures of how many participants are needed in a study. Most quantitative studies should use a sample size calculation to work out how many participants are required to adequately 'power' the study. Intuitively, we assume that the greater the proportion of the whole population studied, the closer we will get to obtaining the true answer for that population. If we do not include enough participants, we are at risk of a type II error, which essentially means that we might miss a statistically significant difference. But how many participants do we need to study to get sufficiently close to the right answer? To answer this you need to know the following:

- the size of the effect (i.e. mean difference and SMD) that is important or meaningful – sometimes this is just estimated, because the desired effect is unknown
- how certain we want to be of avoiding a type I error (i.e. critical level of significance, α)
- the precision and variance of measurements within any sample

The power of a study is the probability that the study will detect a predetermined difference in measurement between the two groups if such a difference truly exists, given a pre-set value of the variance and sample size. You will find most studies report a power of 80–90%; that is, if a difference truly exists between interventions then we will find it on 80–90% of occasions.

Putting it all together – interpreting p-values from a research article

In your role in health care, you will need to be able to distinguish statistical significance (particularly using p values) from practical or medical relevance and clinically worthwhile effects. Rather than focusing on whether the result is statistically significant, you must pay attention to the size of the difference and the CI, because these are what matter for successful practice.

Power and sample size calculations reduce the risk of a study being 'underpowered' because it has too few participants. Therefore, such calculations reduce the risk of a type II error (i.e. missing a statistically significant difference between groups when one exists).

To put this into practice, find a research article that interests you and answer the following questions:

- Can you identify the null and alternative hypotheses? If the article does not state a **hypothesis**, can you try to formulate one for the article?
- Can you explain in simple terms what result, in terms of the **p-value**, would support the **null hypothesis** and what would support the alternative hypothesis (i.e. what is the critical level of significance)?
- What was the desired effect size (i.e. mean difference and SMD)?
- How many participants were required to detect a significant difference?
- What was the probability of obtaining the desired effect size with the number of participants studied?

Now select one statistical finding from the research article and answer the following questions:

- What is the p-value for the statistical test?
- Is this p-value less that the critical level of significance?
- Does the study accept or reject the null hypothesis?
- Did the study perform power and sample size calculations?

Further reading

Salkind, N. J. (2016). Statistics for people who (think they) hate statistics. Sage Publications.

Chapter 12

Introduction to qualitative research design

Introduction and learning outcomes

Chapter 1 introduced quantitative research, which is often known as the 'science of numbers', because it involves researchers testing hypotheses, and using statistics and large-scale data to demonstrate and measure associations between different **variables**. This chapter introduces **qualitative research**, which relies much more on the stories people tell us about an experience or about their perception of something, to help us develop our understanding. Qualitative research involves investigating individuals and groups in their social settings, using several different types of **methods**. The focus is on collecting in-depth information from participants and the goal is to understand rather than to explain.

In Chapter 5 we hypothesised that health sciences students are more empathetic than the general population, and we looked at how we could investigate that **hypothesis** using a quantitative approach. However, we could also investigate that hypothesis using a qualitative research approach. To do this, we would use a smaller number of health sciences students than with the quantitative approach, and would perhaps interview them in-depth about their reasons for choosing a particular health sciences course. This would allow us to develop detailed understandings of the motivations behind the students' choice of a particular discipline. Thus, rather than producing findings in the form of numbers or statistics, we would have data in the form of stories and quotes. Although quantitative and qualitative research methods differ, they are equally valid in health and social science research. Both approaches attempt to understand social reality, but they use different methods to do so.

In the following video Professor Pranee Liamputtong provides an introduction to qualitative research in the context of women's health:

https://doi.org/10.26181/5c11900049cd9

KEY LEARNING OUTCOME

Examine the defining features, strengths and limitations of **qualitative research** and analysis.

ENABLING OUTCOMES

Once you have completed this chapter, you should be able to:

- describe the principles underpinning qualitative approaches
- identify common qualitative approaches and **methods**
- match research questions with appropriate methods of qualitative research
- explain criteria for evaluating qualitative research
- describe the organisation and structure of a qualitative journal article

Principles underpinning qualitative approaches

Qualitative research is underpinned by several principles. The first principle relates to the overall approach. Employing a 'bottom up' approach, qualitative research uses **inductive reasoning** – beginning with the identification of a pattern in the collection of stories or considerations – to help build a theory. Thus, instead of starting with theories, qualitative researchers prefer to work the other way around, entering the research with few preconceived ideas. The key task is to draw out the meanings, perceptions and understandings that individuals and groups attach to behaviours, experiences and social **phenomena**. Hence, the term 'the lived experience' is often used in qualitative research.

Chapter 7 introduced one of the key principles of quantitative research: *objectivity*. The notion of objectivity is rejected in qualitative research, for three main reasons. The first is that little of the quantitative and qualitative research conducted is totally value free, and researchers often enter the field with preconceived ideas. The second is that **objective** reality is based on people's definitions – essentially, this means that we all develop **subjective** meanings of an experience. The third is related to the fact that understandings about issues, experiences and perceptions are **socially constructed**; that is, they are influenced by the world around us. Thus, qualitative researchers acknowledge their own values and biases, and how these might influence the research process.

In investigating and reporting on these experiences and perceptions, qualitative researchers are primarily interested in '<u>thick description</u>' (i.e. a description that includes both the behaviour and the context) rather than explanation. The data are reported in words (primarily the participant's words or pictures) rather than numbers, and researchers use stories and quotes to develop key themes to build theory.

Understanding these experiences is critical in health care. As Pranee Liamputtong explains in her video, qualitative data may often be dismissed as 'soft' in health sciences research. However, certain research questions are best answered using a qualitative approach. For example, if we want to explore community palliative care, and we are interested in developing a detailed understanding of experiences and perceptions, a qualitative approach is the best way to obtain this data, which could then be used to assist in improving service delivery. To exemplify this approach, research case study 8.1 is an abstract from a qualitative study that explores the experiences of caregivers providing rural palliative care.

Research case study 12.1 A qualitative study exploring preparedness for palliative care in rural areas.

The care of people with life limiting illness is increasingly moving away from the acute setting into the community. Thus, the caregiver role is growing in significance and complexity. The importance of preparing and supporting family carers is well established, however less is known about the impact of rurality on preparedness and how preparedness shapes the caregiving continuum inclusive of bereavement. The aim of this 2017 study was to explore how bereaved rural family palliative carers described their preparedness for caregiving. Interpretative phenomenological analysis was employed following semi-structured interviews with four women and six men (N=10, aged 55-87. The experiences of caregivers illuminated a lack of preparedness for the role and were characterised by four major themes: Into the unknown, Into the battle, Into the void and Into the good. The unknown was associated with a lack of knowledge and skills, fear, prognostic communication, exclusion, emotional distress and grief experience. Battles were experienced in a number of ways: intrapsychically (existing within the mind), through role conflict and identity; interpersonally with the patient, clinician and family; and systematically (against health, financial and legal systems). The void was felt during isolation in caregiving, in relinquishing the role, in bereavement and in feeling abandoned by service providers. Positive experiences, such as being valued, included and connected to supports, and the fostering of closer relationships and deeper meaning, occurred less frequently but temporarily buffered against negative aspects. Implications from this study for policy and practice centre on frequent, purposeful & genuine engagement of caregivers. Services and clinicians are encouraged to enhance communication practices, promote meaningful inclusion, address access issues and enhance support at role relinquishment.

Mason, N. & Hodgkin, S. (2018). Preparedness for Caregiving: A Phenomenological study of the experiences of rural Australian family palliative carers, Health & Social Care in the Community (in press).

Common qualitative approaches and methods

Qualitative research methods allow for exploration of multiple realities, focusing on the understanding on how meaning is constructed. In contrast to quantitative methodologies researchers who use qualitative approaches are viewed as co-constructors in the research. Although qualitative research does not produce 'hard figures', it does have rigorous and explicit methodologies for defining problems, collecting and analysing evidence, and formulating and developing theories.

Thus, qualitative methods can be powerful tools for understanding the social and cultural dynamics that people are enmeshed in; such dynamics may not be immediately discerned by structured observation, clinical trials or by **survey** methods. This section explores some of the approaches and data collection methods associated with qualitative research.

APPROACHES TO QUALITATIVE RESEARCH

This section outlines four approaches to qualitative research: narrative enquiry, phenomenological study, **ethnographic** research and **grounded theory**. It is stressed that this is not an exhaustive list as there are numerous other qualitative approaches. Here, the more common approaches are described.

Narrative enquiry

Narrative enquiry is concerned with stories or accounts of an event, as described by the person involved in that event. The purpose is to discover what meanings individuals give to particular life episodes. The approach begins by looking at stories told by individuals; these stories are examined closely for meaning, plot and metaphor, to understand how people make sense of their stories. There are many forms of narrative analysis and it is beyond the scope of this topic to cover them all. However, in broad terms, the researcher examines how stories compare and contrast in terms of plot, characters, metaphors and interpretations.

Phenomenological study

In contrast to narrative enquiry, a phenomenological study describes the meaning of the lived experiences for a group of individuals about a concept or a **phenomenon**. **Phenomenology** is probably the most generic out of all the qualitative research designs and probably one that is the most used by qualitative researchers, particularly those conducting health research. The purpose of phenomenological approach is to illuminate the specific – to identify **phenomena** through how it is perceived by the actors in a situation. Such an approach would consider the following: What is the meaning,

structure, and essence of the lived experience of this **phenomenon** for **this person or group of people** in particular, as there are multiple ways of interpreting the same experience. It is based upon experiences occurring in everyday life, in natural settings, by persons from all walks of life.

So, for example, what meanings did people construct about their experiences of being earthquake and tsunami survivors in Indonesia in September 2018? You want to capture the **subjective** nature of the phenomenon. What was it really like? In order to really get that in-depth understanding you need to have a way to gain an in-depth exposure to peoples' experiences and interaction with the phenomenon of interest. This is where qualitative methods come in. In a phenomenological study, the interviewer starts with one question, **"tell me about the time when you experienced the tsunami "** and considers the participants responses to be the prompts to peel away like the layers of an onion to get to the essence of the experience. So the assumption is that there is always a core, a nature of things, a shared understanding amongst a group of people, an essence that holds a phenomenon or experiences together.

Ethnographic research

In **ethnographic** research, researchers spend considerable time observing a group of people, their cultures and rituals. As a process, ethnography involves prolonged observation of the group, typically through participant observation; for example, the researcher may become immersed in the social setting over an extended period of time. The researcher also listens to and engages in conversations, and might also collect information through in-depth interviews and by gathering documents. An interesting development in **qualitative research** in recent years has been the growth of interest in visual materials. Visual ethnography has become increasingly popular, both as sources of data and as prompts for discussion by research participants. Research case study 8.2 is an example of a visual ethnography study.

Research case study 12.2 Example of a visual ethnography study

Radley and Taylor (2003) were interested in the role that the physical setting of a hospital plays in patients' recovery. Nine patients in a ward were asked to take photographs on the ward some days after their surgery or medical investigation. Each patient was supplied with a camera and asked to take up to 12 photos of things that were significant to them in the hospital. The researchers stayed with the patients while they took their photographs. Patients were interviewed the following day and then a month later in their own homes. On each occasion, patients were asked about all the photographs and which ones best expressed their stay in hospital.

Radley, A., & Taylor, D. (2003). Images of recovery: A photo-elicitation study on the hospital ward. Qualitative Health Research, 13(1), 77–99.

Grounded theory

Grounded theory differs from other qualitative methodologies because of its focus on generating data from which to develop a theory. The origins of grounded theory can be traced back to the work of Barney Glaser and Anselm Strauss during the 1960s, when quantitative research was considered to have more significance and relevance to studying sociological events. Glaser and Strauss (1967) challenged this thinking, by developing an approach to **qualitative research** that focused on the development of theory. Their ground-breaking study on the experience of dying led to the creation of a research approach that detailed how theory was grounded in and developed from the data. They created a comprehensive systematic approach of data collection, data coding and analysis, with the aim of moving beyond descriptions of the **phenomena** and instead generating a new understanding of the phenomena; that is, a grounded theory. Grounded theory approaches are particularly useful in qualitative research in cases where little is known about a topic.

Glaser, B. & Strauss, A. (1967). The discovery of grounded theory strategies for qualitative arch. Chicago: Aldine Pub. Co.

DATA COLLECTION

This section describes the three main **methods** of data collection: individual in-depth interview, **focus group** and clinical data mining.

Individual in-depth interview

The individual in-depth interview represents the most common form of data collection in qualitative research. In an in-depth interview, the researcher talks to the participant about the research, and invites the participant to talk about their life or their perspectives on a particular topic. Thus, there is greater interest in the *participant's* point of view, and the interviewer follows up on the key issues identified by the participant, rather than the other way around. Research case study 8.3 provides an excerpt **from a New Zealand** study in which the authors interviewed 17 street-based sex workers who had entered the industry under the age of 18. The excerpt illustrates how authors use quotes from participants when they discuss their findings.

Research case study 12.3 Example of a study using individual in-depth interviews

Many of the street based workers who participated in the in-depth interviews ran away from home at an early age and were living on the street prior to starting sex work ...The street based participants in this study had not experienced stable and supportive family lives and some had experienced foster care.

'Well since I was 11, I was put into CYF's (Child, Youth and Family) care... And I learned to deal with the fact that's my family, because my family hasn't been

around me. You know, my mum walked out on me when I was two and a half, and my dad, he's just an idiot'.

Few-street based workers discussed how long they would remain in the sex industry. Street-based workers who discuss this described leaving when they could 'turn their life around', such as when they could get into a drug rehabilitation program.

'So then (on entering a drug rehabilitation program) I will be quitting because I am just at that point where I need to turn my life around, and I know I can do it, because I'm just, you know, I've had enough and I want to do it. I want to make changes, I can't do that if I'm still working you know'.

Abel, G. M., & Fitzgerald, L. J. (2008). On a fast-track into adulthood: an exploration of transitions into adulthood for street-based sex workers in New Zealand. *Journal of Youth Studies*, *11*(4), 361–376. doi:10.1080/13676260802104808

An in-depth interview can be totally unstructured, semi-structured or standardised, although in all cases the interview process is flexible. In an unstructured interview, the researcher uses an informal conversational tone, and has a brief set of prompts concerning the topic. The researcher invites the participants to respond to some open-ended questions; for example, 'I am really interested in knowing more about the experience of living with your illness. Would you mind taking me through a day in your life?' In a semi-structured interview, the researcher has a list of pre-set questions, but makes allowance for the participant to expand upon responses; for example, 'You said earlier that you had had some negative experiences with the hospital. Would you mind telling me more about that?' In a standardised open-ended interview, all participants are asked the same questions in a structured way.

Focus group

A focus group is a form of in-depth interviewing that is conducted with a group of people rather than an individual. Focus groups have been used extensively in consumer research and are rapidly gaining popularity as a research methodology in health research. As the name implies, in this approach a researcher brings a group of people together (usually 6–8 people) for a focused in-depth discussion on a particular topic. Usually, participants in a focus group share a common interest or experience. The aim is to generate collective and differing perspectives on a particular topic, informed by the interaction of participants. Thus, the interest is on a collective discussion of a topic as opposed to an individual one.

'Six women standing and sitting' from Pexels used under Pexels Photo Licence

The researcher's role is to guide the discussion. Individuals in a group will often challenge each other's views. It requires a skilled moderator to ensure that all viewpoints are heard, that tensions are skilfully dealt with and that the focus remains on the topic to be explored. Hence, **focus groups** require careful organising and consideration of group dynamics. Some research topics may be too sensitive to explore in a group, particularly if personal information is shared. Confidentiality in focus groups cannot be ensured, because participants may repeat what is said in a group.

Clinical data mining

Clinical data mining involves extracting and analysing data to uncover hidden patterns. It is a form of practice-based research that uses secondary data analysis (i.e. analysis of data collected for another purpose), as illustrated by Research case study 8.4. In health sciences, we collect and record large amounts of information in relation to our clients or patients. We collect data on their problems, and the various interventions that have been used or promoted. These data can be used in EBP to answer questions such as:

- What are the characteristics of the clients or patients we see?
- What are we actually doing for them?
- What seems to work?
- What produces negative outcomes?

Research case study 12.4 Example of clinical data mining

A medical consultant working in the Diabetes Unit at the Royal Children's Hospital was concerned in regards to the number of adolescents frequently being re-admitted to hospital. After gaining ethics approval, David (social worker) examined medical records, social work files and other case records to identify how many and which patients were experiencing multiple admissions. When reading the files, David took notes on all the psychosocial issues identified and constructed a list of issues they had in common. He was able to identify some common themes related to re-admissions such as the incidence of learning disorders, parental psychological problems and family dysfunction. As a result, he was able to suggest some avenues for intervention that included patient-focused, family-relationship and parent-focused interventions.

Nillson, D. (2002). Psycho-social problems faced by "frequent flyers" in a pediatric diabetes unit. *Social Work in Health Care*, *33*(3–4), 53–69. doi:10.1300/J010v33n03_05

Match research questions with appropriate methods of qualitative research

Chapter 5 discussed research design and choosing a design that best suits the research question. In choosing a design, the researcher needs to be aware of the advantages and disadvantages inherent in both quantitative and qualitative designs. Often, researchers will choose designs that they feel most comfortable with, rather than establishing first how to best answer a research question. Rather than thinking, 'I am no good with numbers' or 'I do not understand **qualitative research**', a better approach is for researchers to look at the research question and ask 'What is the best way to find answers to my research question and, more specifically, am I theory testing or am I theory generating?'

Methodological quality

As was highlighted in Chapter 7, in relation to quantitative research, you need to become skilled at examining the strength of the evidence you are reviewing. The same is true for qualitative research.

Peer-reviewed journal articles reporting on qualitative research tend to be longer than those reporting on quantitative research. In examining qualitative research, things to look for are whether the researchers have provided a detailed description of:

- the research site
- how the sample was obtained
- how many participants were in the study
- how the data were analysed
- the actual data (e.g. in the form of quotes and case examples)
- positionality of the researcher

You should also assess the article in terms of whether the researcher has established that the data are reliable, credible and trustworthy (discussed further in Chapter 13).

The organisational structure of a qualitative journal article

Chapter 5 presented the basic structure and organisation of a quantitative journal article, noting that you need to be familiar with this before you start reading and appraising quantitative research. The same applies to **qualitative research**.

A key difference between a quantitative journal article and a qualitative journal article is that, in the latter, the language is not as **objective** or formal, and often the researcher might write in the first person. Another feature of a qualitative journal article is that data are presented in the form of direct quotes or case examples, rather than tables; however, figures and diagrams are sometimes included.

A typical journal article dealing with qualitative research has the following structure (See **Annex 2**):

Title – Informative, attract the reader's attention, should accurately reflect the nature and focus of the study.

Abstract – Short summary, provides an overview of what the research is about, what was done, how it was done, what was found, and what the results mean.

Keywords – 6-8 keywords used to draw the reader's attention, also used to locate articles in electronic databases.

Introduction – Brief overview of previous relevant research, rationale, provides a rationale for the study and an outline for what the research is aiming to do. Authors highlight a gap in knowledge and describe what their study will provide in relation to this gap. In qualitative journal articles, the theory or framework used may be introduced.

Methods – Discusses the design of the research and method employed. Explanation of why a qualitative approach has been used, including discussion of ontology and epistemology. Methods used to ensure rigour are discussed. Detailed description of research site, population, participants, and the process of data collection and analysis.

Findings – Findings (results) can be presented as themes, verbatim quotations are used to elaborate on the explanation of the findings. Findings can be presented in different ways e.g. given without interpretation, or given with interpretations but a detailed discussion is left to the discussion section, or findings and discussion can be combined.

Discussion – Summarises and interprets findings, relates the findings back to previous research, literature, and theoretical framework. Considers

the original research question or hypothesis, and discusses the clinical implications for the client and the profession.

Conclusion – Provides any limitations of the research and recommendations for future research.

Further reading

Liamputtong, P. (2013). Qualitative research methods (4th ed.). South Melbourne: Oxford.

Denzin, N., & Lincoln, Yvonna S. (2017). The Sage handbook of qualitative research (Fifth ed.). Los Angeles: Sage.

Higginbottom, G., & Liamputtong, P. (2015). Participatory qualitative research methodologies in health. Los Angeles: Sage.

Liamputtong, P. (2011). Focus group methodology: Principles and practices (1st ed.). Thousand Oaks, CA: Sage Publications.

Chapter 13

The role of rigour in qualitative research

Introduction and learning outcomes

Chapter 6 noted that establishing the **validity** of quantitative research requires strict observation of a number of procedures, to ensure that the research remains free of **bias**. Those procedures are not applicable to **qualitative research**.

The aim of qualitative research is to provide a rich description of the meanings people give to their experiences. Hence, such research is not interested in measurement validity, or in threats to internal and external validity, because it is not important for qualitative research to establish a causal link between two or more **variables**, or to generalise findings to the broader population. Instead, qualitative researchers would argue that qualitative research has its own validity, because researchers become immersed in the real world of their participants, rather than attempting to measure something through artificial means.

Qualitative researchers prefer to use the term 'rigour' rather than 'validity' and '**reliability**'. Qualitative research has established rigorous **methods** to ensure faithful representation of the stories and experiences of participants. For example, qualitative researchers keep detailed field notes; also, where possible they use audio and video recordings to keep accurate records of data. They also send transcripts of interview or focus data back to participants, to ensure that the participants have an accurate record of any interviews.

When you are reviewing journal articles as part of your role in health care, you will find yourself wondering whether the results of the study are accurate and trustworthy. If the results are neither accurate nor trustworthy, then you need to consider how this may affect the way you undertake your work or practice (i.e. how you assess and manage clients). This chapter provides detail about how qualitative researchers ensure that their research is rigorous and trustworthy.

KEY LEARNING OUTCOME

Critically appraise a journal article and note several considerations relating to rigour in qualitative studies.

ENABLING OUTCOMES

Once you have completed this chapter, you should be able to:

- critique a qualitative journal article, and identify issues relating to trustworthiness, credibility, transferability, dependability and confirmability

Rigour and trustworthiness

You can judge the rigour and trustworthiness of a **qualitative research** article by considering the following key criteria: *credibility, transferability, dependability* and *confirmability.* Explanations and discussion of these terms are presented below in the context of qualitative research conducted by Dr Mandy Ruddock-Hudson (La Trobe University) in her exploration of the psychological reactions to injury in the Australian Football League (AFL). In the following video, Dr Ruddock-Hudson discusses some of the issues of rigour in her qualitative research with AFL injuries:

https://doi.org/10.26181/5c11900986795

You can also get an overview of Dr Ruddock-Hudson's research in the following journal article:

Ruddock-Hudson, M., O'Halloran, P., & Murphy, G. (2012). Exploring psychological reactions to injury in the Australian Football League (AFL). *Journal of Applied Sport Psychology*, *24*(4), 375–390.

CREDIBILITY

Credibility is an evaluation of whether the research findings represent a 'credible' conceptual interpretation of the data drawn from the participants' original data; that is, whether the findings are truthful and believable. You need to consider whether the descriptions provided by the participants in the study are represented, and whether they fit with the explanation from the researcher or author of the paper.

In the study by Ruddock-Hudson et al. (2012), the researchers used a technique called 'member checking' to enhance the credibility of their work. This technique involved evaluating the accuracy of transcription:

> *One quarter of the sample was randomly selected and sent copies of the transcripts to review, to ensure that the interviews had reflected what they had said.*

The article also included several quotes from transcripts that exemplified themes. For example, this quote from a player exemplified the theme *'The influence of social support'*.

> *My family has always been there to provide me with support, even if I wasn't a footballer, they will always be there no matter what. The medical staff were great when I was injured. They were constantly keeping me up to speed on how my injury was progressing and always asking how I was doing.*

Ruddock-Hudson, M., O'Halloran, P., & Murphy, G. (2012). Exploring psychological reactions to injury in the Australian Football League (AFL). *Journal of Applied Sport Psychology*, *24*(4), 375–390.

Another way to enhance credibility is to use triangulation, whereby a study uses at least three methods, with the aim of double or triple checking the results. For example, the themes generated from the transcripts of the interviews can be crosschecked with the results of psychometric tests (i.e. measurement of knowledge, abilities, attitudes and personality traits). Triangulation was not performed for the Ruddock-Hudson study, and this is noted in the article as a limitation of the study.

TRANSFERABILITY

***Transferability* is a characteristic that indicates to what extent the study findings can be generalised or applied to other individuals or groups, contexts or settings.**

When reviewing Dr Ruddock-Hudson's study of injuries in AFL players, you might be interested in whether her findings can be applied to other groups (e.g. elite female netballers), or different contexts or settings (e.g. the workplace). Transferability indicates the degree to which qualitative findings inform and facilitate insights within other contexts, other than that in which the research was conducted.

In the limitations section of the article Dr Ruddock-Hudson discusses the fact that, because the sample was obtained from only one club, 'the club culture … may have systematically influenced the results'. This indicates that the results of her report may be limited in their transferability to other AFL clubs.

DEPENDABILITY AND CONFIRMABILITY

***Dependability* is an assessment of the quality of the integrated processes of data collection, data analysis and theory generation, whereas *confirmability* is a measure of how well the inquiry's findings are supported by the data**

collected. Sometimes, research articles report a process whereby a competent peer undertakes an independent audit to examine original transcripts, data analysis documents, field journals, comments from the member checking and the text of the journal manuscript. As part of this process, the auditor can evaluate the degree and significance of researcher influence – to what extent the researcher has influenced or biased the interviews.

In the following video Associate Professor Kath Ryan discusses some of the issues of rigour in her **qualitative research** on breastfeeding:

https://doi.org/10.26181/5c11903056cf3

Further reading

Ruddock-Hudson, M., O'Halloran, P., & Murphy, G. (2012). Exploring Psychological Reactions to Injury in the Australian Football League (AFL). Journal of Applied Sport Psychology, 24, 375-390.

Baillie, L. (2015). Promoting and evaluating scientific rigour in qualitative research. Nursing Standard (Royal College of Nursing (Great Britain): 1987), 29(46), 36-42.

Qualitative data analysis (2013) p.181. In. S. Polgar & S.A. Thomas, Introduction to research in the health sciences, 6th Ed., Churchill Livingstone Elsevier.

Chapter 14

Introduction to mixed methods designs

Introduction and learning outcomes

Previous chapters looked separately at quantitative and **qualitative research**. They discussed research design and choosing a design that best suits the research question, and they noted the need to be aware of the advantages and disadvantages inherent in both quantitative and qualitative designs.

The best way to answer a research question may be through a range of data collection **methods**. Indeed, several key writers on research methodology argue that research at any point in time falls within a research cycle that moves from inductive to deductive logic. The decision to use qualitative or **quantitative methods**, or indeed both, depends on two things: the research question, and the phase of the research cycle. Qualitative methods provide for richly textured data, whereas quantitative data, through the development of sophisticated measurement tools, can provide strong evidence of patterns. In health-care studies, a mixed methods approach that integrates both quantitative and qualitative data often provides a more holistic understanding of the research issue.

Mixing methods involves collecting data either simultaneously or sequentially; it involves gathering both numeric information as well as textual information, and the final database represents both quantitative and qualitative information. Several key texts discuss the history of mixed methods and its application in research (e.g. Teddlie & Tashakkori, 2009; Creswell & Plano Clark, 2011). There are also a range of journals committed to the publication of mixed methods designs (e.g. *Journal of Mixed Methods Research* and *Quality and Quantity*).

KEY LEARNING OUTCOME

Match mixed methods designs to key research questions.

ENABLING OUTCOMES

Once you have completed this chapter, you should be able to:

- describe the reasons underpinning mixed methods approaches
- identify four common mixed methods designs

In the following video Dr Suzanne Hodgkin discusses principles of mixed methods research:

https://doi.org/10.26181/5c118fe62df0b

Reasons for using mixed methods

Back in the late 1980s, several publications describing and defining mixed methods research began to appear across disciplines. For example, Greene, Caracelli & Graham (1989), identified the following five reasons for conducting mixed methods studies:

To triangulate the data – collecting both quantitative and qualitative data simultaneously, comparing and contrasting the findings and merging them.

To complement the data – sometimes the researcher is seeking elaboration, enhancement, illustration and clarification from mixing the results of one method with those of another.

To develop the data – when the researcher uses the results of the first method to develop or inform research using the other method.

To understand paradoxes or contradictions emerging from one type of data – when the researcher follows up on findings that need further explanation.

To expand the data – when the researcher seeks to extend the breadth and range of a study by using different methods for different components of the study.

Types of mixed methods designs

This section discusses the main types of mixed methods designs: convergent parallel design, explanatory sequential design, exploratory sequential design and embedded design.

CONVERGENT PARALLEL DESIGN

Convergent parallel design is used to obtain different but complementary data on a topic. The purpose is to combine the strengths of each type of data; thus, both quantitative and qualitative data are collected simultaneously. This type of study might include cross-sectional data in the form of a **survey**, **focus group** data and in-depth interviews all occurring at the same time, with the results of each then being brought together. For example, a local community health centre might be interested in conducting a health-care needs study, and might combine several different types of data (e.g. questionnaires, in-depth interviews and focus groups) in the study. This type of design is illustrated in Figure 14.1.

Figure 14.1 Convergent parallel design

EXPLANATORY SEQUENTIAL DESIGN

Explanatory sequential design is a two-phased study. The first phase involves the collection of quantitative data that specifically addresses the study's questions. In the second phase, qualitative data might be obtained to help explain or build on the initial quantitative results. An example is a study on the retention of older health-care workers (Hodgkin et. al 2017). In Australia, a shortage of staff across a number of health sector positions presents a challenge, and this problem will be compounded by the likely retirement of a large number of baby-boomers in the next 10 years. Exploring factors that serve as barriers and incentives to keeping this **cohort** in the workforce is critical. Thus, the overall research question for this study was: 'What are the organisational and social factors that impact on the retirement intentions of health care workers who are aged 55 years and over? In the first phase, participants (n=299) completed a **survey** that contained the following measures: demographic **variables**, retirement intentions, an effort–reward imbalance measure and a general health measure. In the second phase, a

smaller sample (n=17) participated in an in-depth interview that explored both retention and retirement intentions, to help explain some of the quantitative findings. The process of explanatory sequential design is shown in Figure 14.2.

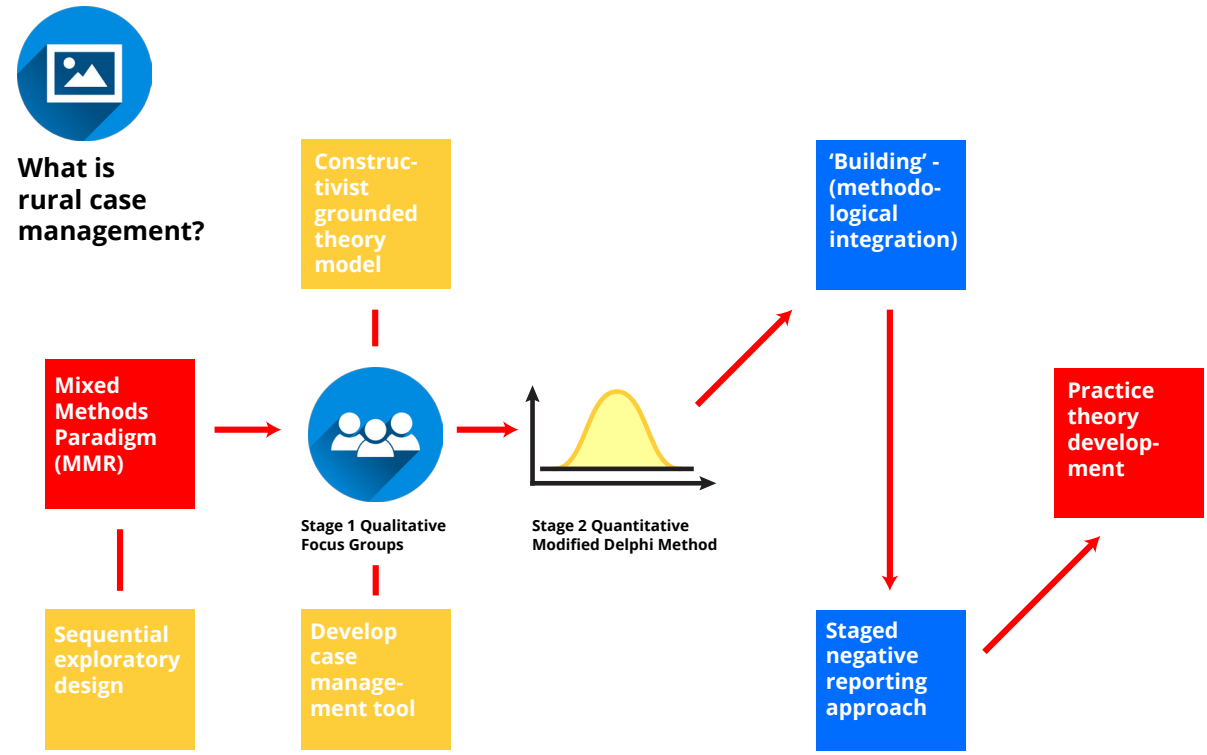
Figure 14.2 Explanatory sequential design

EXPLORATORY SEQUENTIAL DESIGN

In contrast to explanatory sequential design, an exploratory sequential design process begins with and prioritises qualitative data. In this type of design, the results of the first stage (qualitative) are used to develop the second stage (quantitative). The design is based on the premise that exploration of an issue or concept is required first. This type of design is useful in developing theories or concepts when measures or instruments are not available. A recent study (Dellemain, Hodgkin & Warburton, 2017) used exploratory sequential design to develop a practice theory for rural case management. The authors argued that although a theory on case management has been developed, there has been little research into the impact of rurality on this type of community work, particularly for the Australian context. The design selected was a qualitative dominant, sequential, exploratory mixed method design, the aim of which was to develop community-based rural case management practice theory. Figure 14.3 shows a diagram of exploratory sequential design, and Figure 14.4 illustrates a study on rural case management.

Figure 14.3 Exploratory sequential design

Figure 14.4 Use of exploratory sequential design to develop a practice theory for rural case management



'Icon landscape' from Pixabay used under CC0 1.0
'Icon teamwork' from Pixabay used under CC0 1.0

EMBEDDED DESIGN

Often, in health research, one dataset alone cannot fully explain the existence of patterns in the data. In an embedded design, the use of a mixed method approach can offset these limitations by adding different types of data that provide a supportive secondary role. One example is a study of the workflow and work patterns of Australian residential aged care facilities (Hodgkin, Warburton & Savy, 2012). This research was concerned with accurately reporting and documenting the activities undertaken by the health-care workforce (e.g. division 1 nurse, division 2 nurse, allied health practitioner and ward clerk). The research used a structured observation technique over a two-hour period, to document each role and the time taken to do each activity. This quantitative data was supplemented by qualitative data in the form of structured interviews with key personnel. This approach provided crucial contextual data to help explain contextual factors; for example, whether there were staff shortages for particular roles, the qualifications held by staff in particular roles, and how the layout of the facility affected a task. Figure 14.5 shows a diagram of an embedded design.

Figure 14.5 Example of an embedded design

In the following video Dr Jo Rayner discusses how quantitative and qualitative data can be combined in the one study (i.e. mixed methods) in the context of The Tall Girls Study:

https://doi.org/10.26181/5c11901881aa6

Further reading

Creswell, J. W., & Clark, V. L. P. (2017). Designing and conducting mixed methods research. Los Angeles: Sage.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. Educational evaluation and policy analysis, 11(3), 255–274. doi:10.2307/1163620.

Hodgkin, S. (2008) Telling it all: A Story of Women's Social Capital Using a Mixed Method Approach. Journal of Mixed Methods Research 2(4), 296-317.

Hodgkin, S., Warburton, J., & Savy, P. (2012). Using mixed methods to develop and implement a work sampling tool in residential aged care. International Journal of Multiple Research Approaches, 6(1), 23–32. doi:10.5172/mra.2012.6.1.23

Teddlie, C., & Tashakkori, A. (2009). Foundations of mixed methods research: integrating quantitative and qualitative approaches in the social and behavioral sciences. Thousand Oaks: Sage.

Chapter 15

Ethics in research and practice

Introduction and learning outcomes

This chapter provides an overview of key concepts and issues that underpin ethical research. In particular, we will consider why **ethics** is an issue in research, when ethical review is required, and what roles ethics committees play in ensuring the protection of human rights. We will also consider the key principles of ethical research, and examples of ethical and unethical research.

When we think of ethics, we are usually concerned with moral standards that govern behaviour. When conducting research, we need to ensure that ethical principles and values always govern research involving people. Researchers in health and human services rely on members of the public to participate, and it is imperative that participants' human rights are protected, and that ultimately we do no harm.

'Pills' by Michal Jarmoluk from Stocksnap used under CC0 1.0

It is always important to consider the following questions:

How should we treat people on whom we conduct research?

What are the rules of conduct?

How can we be sure that researchers are acting in a moral or ethical way?

Are there activities that we should or should not engage in for research?

How do we identify and overcome issues of exploitation?

KEY LEARNING OUTCOME

Explain how common ethical issues can influence health research and practice.

ENABLING OUTCOMES

Once you have completed this chapter, you should be able to:

- discuss the roles of ethics committees
- explain the terms 'informed consent', 'beneficence', 'vulnerability', 'justice' and 'contribution to knowledge'
- analyse examples that demonstrate ethical issues in health research and practice
- recommend how EBP might be ethically improved.

The role of ethics committees

In the past, there were several examples of research that were considered to be unethical and in violation of human rights. Examples of unethical research can be traced back to World War 2. During and before World War 2, information emerged about abuses perpetrated by physicians and scientists (in the name of **experimental** research) on people in Nazi concentration camps; for example, subjecting people to freezing. There was subsequent acknowledgment that unethical research was sometimes found in many other settings. This led to the development of the *Nuremberg Code: Directives for human experimentation* (1949).

Other key events in developing key principles for ethical research include the World Medical Assembly's *Declaration of Helsinki* (1964) and the Australian National Health and Medical Research Council's (NHMRC's) first statement on human experimentation in 1966. Although these key principles were developed, their application was left to researchers. Reports of abuses led to the development of human research ethics committees (HRECs); the essential brief of such committees was to review all applications for proposed research, to ensure that ethical principles of research were upheld. To illustrate, in 1973 the NHMRC made it a condition that all grant applications must have received ethical approval from an HREC.

In the following video Dr Deirdre Fetherstonhaugh discusses the role of human ethics committees:

https://doi.org/10.26181/5c11902a1350e

Since 1985, all human research conducted by government organisations, universities and health authorities that involves potential risk or inconvenience to people must be approved by an HREC. Such committees usually comprise researchers, health and human services professionals, a lawyer and at least one lay member of the public. HRECs play an important role in the Australian and international systems of research review, protecting the public from unethical research.

When submitting an Ethics Application, the researcher is required to submit a research proposal to the HREC. This ensures that the needs of potentially vulnerable people are taken into consideration. Research that involves negligible risk (e.g. the use of existing data sources that contain non-identifying information) may be exempted from a submission to an HREC.

For further information: La Trobe University Research Ethics Application form and guidelines.

Explanation of informed consent, beneficence, vulnerability, justice and contribution to knowledge

Today, it is generally accepted that research involving humans must meet the following ethical criteria: informed consent, beneficence, vulnerability, justice and contribution to knowledge. Although these areas overlap, they provide a useful guide in the conduction of ethical research.

INFORMED CONSENT

The principle of informed consent means that participants should be given as much information as possible to make an informed decision about their participation in a study. All risks involved in the research must be explained, alongside the possible benefits. If procedures are to be used, a clear and honest explanation of these must be given.

The decision to participate should be made without any formal or informal coercion. Researchers should be attuned to whether participants are limited in their ability to fully understand what their participation might involve. To ensure this, participants are usually required to sign an information sheet written in plain language that summarises key aspects of the research, and outlines what their participation involves and any potential risks arising, what will happen with the data collected, and whether participants can

access findings and results. An example of an information sheet provided to a potential participant, which illustrates all the information that should be given to a participant before they agree to being involved in a study, is given in **Annex 1**.

There are examples in research where informed consent cannot be guaranteed. In **ethnographic** studies sometimes the participants' behaviour is observed prior to obtaining consent. The reason for doing so relates to the argument that participants often change their behaviour if they know they are being observed. It is recommended that informed consent is obtained post hoc in these cases.

BENEFICENCE

The principle of beneficence requires researchers to assess and take into account the risk of harm to participants. The potential for possible harm must be identified and minimised. Harm could encompass physical harm, harm to participants' development, and psychological or emotional discomfort or distress. It is the researcher's responsibility to anticipate and guard against potential harm.

The issue of harm also relates to the maintenance of confidential records. Care must be exercised in record keeping of confidential files. Participants have the right to anonymity (i.e. that they will not be able to be identified in any way) and the right to confidentiality (i.e. that information will not be used for any other purpose other than the research). It is vital that participants cannot be identified when findings are published. The most effective way of ensuring anonymity is to collect non-identifying data. This is easier to achieve in quantitative research that in **qualitative research**.

The issues of anonymity and confidentiality are critically important in qualitative research. To guard against participant identification, pseudonyms are used and identifying information is kept in a separate place from the data.

VULNERABILITY

The principle of vulnerability refers to guidelines that specifically protect vulnerable populations. Abuse through past research has informed requirements for additional care in research with vulnerable groups. Such groups include:

- children and young people (especially those on child protection orders)
- people in dependent or unequal relationships (e.g. women who have faced partner violence and older people)
- Aboriginal and Torres Strait Islander people

- people highly dependent on medical care
- people with a mental illness
- people with an intellectual or mental disability
- people who might be involved in illegal activities
- relatives of sick people whose English is not their first language

All such vulnerable groups should be treated with special care, taking into consideration the likely effects of participation, to ensure that they are able to give informed consent and are participating in a voluntary way. In designing research with vulnerable groups, particular care needs to be taken with ensuring informed consent and voluntary participation.

JUSTICE

The principle of justice refers to the ethical value of the research. There should be justice in the distribution of both the burdens and benefits of research, with no single group being over-researched. A key question here is, 'Who benefits from the research?' The benefits of research must flow across society, rather than increasing the advantage of some groups while neglecting the interests of others.

It is important to understand that the power of researchers often contrasts with that of their participants. With the exception of **action research** and some feminist research, the researcher controls how the study is to be conducted, and what to do with the information collected. Certainly, there are benefits for the researcher in conducting research; for example, in the form of prestige, and the publication of findings in reputable journals. It is therefore important to consider issues of exploitation.

CONTRIBUTION TO KNOWLEDGE

The principle of contribution to knowledge refers to whether the research is worthwhile. There should be benefits in the form of contribution to knowledge, and of direct benefits to the participants and to the public good. A community that is the subject of research should benefit from the research, rather than being disadvantaged by it.

Analyse examples that demonstrate ethics issues in health research and practice

There are many examples where researchers have been unethical; for example, researchers may have used deception, or harmed participants in some way. In applying the principles of ethical research, the following questions provide a useful starting point, and may help in deciding whether there are any potential ethical issues with the research:

Has the researcher followed ethical principles?

Would the participants be willing to do further research of this kind?

Would I be happy for members of my own family to participate?

Consider Case study 15.1. Using the questions given above, consider whether the researcher has done anything wrong, and if so, what has the researcher done wrong?

Research Case study 15.1 Humphreys' tearoom trade

In the late 1960s, Humphreys used a covert research methodology to discover the social background of men who engage in homosexual acts. Humphreys' method involved observing homosexual encounters in public places including truck stop restaurants (tearooms). He acted as a 'Watch Queen', warning men when a member of the public was approaching. In this role, he was able to make a note of the men's car registrations and he used this information to track down addresses. He visited the men a year later, under the guise of collecting information for a 'health survey'. During this visit he gained information about the men's socioeconomic background, leading him to find that they were 'ordinary' citizens, with many being married and holding 'respectable' jobs.

Humphreys, L. (1970). Tearoom trade: impersonal sex in public places, (Observations). Chicago: Aldine.

It could be argued that Humphrey's study did not allow informed consent. Instead, the research approach used both disguise and deception. It also involved an invasion of privacy, with records linking car registrations and home addresses of the men. Thus, some participants may have been identified against their will. Many of the participants were married, and being identified as a practising gay man was an invasion of their privacy.

Consider Case study 15.2. Using the questions given above, consider whether the researcher has followed ethical principles and whether there is an issue of

harm to participants in this study.

Research Case study 15.2 Milgram's experiment

A researcher recruited research participants and paid them a small sum to take part in an experiment, which they were told was about memory and learning in different conditions. The participant was introduced to a 'learner' (an actor) and an 'experimenter' (the researcher). The experimenter told the participant that the learner would have to memorise word pairs, and that when they got them wrong a shock would be administered by the participant (no shock was actually administered, but the participant did not know this). The participant was given a small shock to indicate the 'learner's' experience. The participant was told that they would have to raise the shock intensity 15 volts for each incorrect answer. The 'learner' cried out each time a 'shock' was administered, more loudly as the voltage increased. If the participant wanted to stop taking part, the researcher used successive prompts; for example, 'Please continue', 'The experiment requires that you continue', 'It is absolutely essential that you continue', 'You have no other choice' and 'You must go on'

Milgram, S. (1974). Obedience to authority. London: Tavistock.

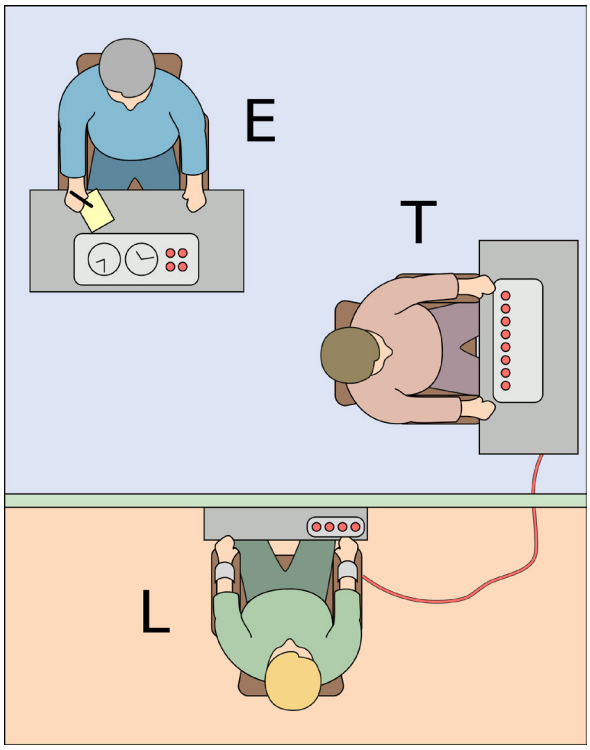


Illustration of the setup of a Milgram experiment. The experimenter (E) convinces the subject ("Teacher" T) to give what he believes are painful electric shocks to another subject, who is actually an actor ("Learner" L). Many subjects continued to give shocks despite pleas of mercy from the actors.

As with the Humphreys' study, Milgram's experiment breaches the key ethical principles discussed earlier. For example, there were issues around deception and confidentiality, with participants not given full information about the nature of the research and the potential harm that may have arisen. In addition, the researcher executed considerable power over the participants. Finally, there were issues around psychological harm, given that many of the participants showed considerable distress and anxiety.

How to ethically improve evidence-based practice

The Humphreys and Milgram cases are famous for transgressing key ethical principles. These cases are thought of as being associated with a particular form of research-disguised observation or experiments involving deception. However, ethical principles can be compromised in all types of research. To ethically improve EBP, the researcher must consider the ethical principles at each stage of the research process. For example, ethical issues are involved in the development of research questions (how compelling is the research?), research design (sample selection) and methodology (what will participation mean?), and in the reporting of results (maintaining confidentiality).

In particular, the following ethical principles must be adhered to:

- research participants must be volunteers and must be recruited in a way that allows them to 'opt in' to a study
- consent to participate in research must be informed
- no harm should follow as a consequence of participating in the research
- sensitive information should be protected – confidentiality is important
- deception should be avoided
- existing databases – such as client records – should not be accessed without approval of those who provided the data.

Some ethical DO NOTS include the following – do not:

- include or continue to research participants demonstrating resistance or discomfort
- attempt to convince people to participate
- fail to explain all relevant aspects of the study *before* they agree to participate
- promise anonymity and confidentiality if this may not be honoured
- use procedures that may entail physical or mental stress
- include techniques of questionable safety
- violate professional research standards
- accept contracted research that violates ethical or professional standards.

EXERCISE 15.1 ETHICAL CONSIDERATIONS FOR A QUALITATIVE STUDY OF ALLIED HEALTH-CARE STAFF

Imagine that you are seeking to undertake a qualitative examination of retention of allied health-care staff in public hospitals in Victoria:

- What are the ethical issues you will face?
- How would you deal with those ethical issues?
- What would you need to include in an information sheet to accompany your application to the La Trobe University HREC?

Further reading

Alston, M., & Bowles, W. (2003). Research for social workers: an introduction to methods. Crows Nest, NSW: Psychology Press.

Bryman, A. (2008). Social research methods (3rd ed.). Oxford: Oxford University Press.

Habibis, D. (2006). Ethics and social research. In M. Walter (Ed.), Social research methods: an Australian perspective (pp. 53–82). South Melbourne: Oxford University Press.

Ramcharan, P. (2010). What is ethical research? In Research methods in health: foundations for evidence-based practice (pp. 27–41). South Melbourne: Oxford University Press.

Annex 1

Example of a participant information statement and consent form

Annex 1

The following is content from John Richards Centre| Research into Aged Care in Rural Communities| College of Science, Health and Engineering | Albury/ Wodonga Campus | La Trobe University.

Resesrch Project Title		
Role	**Name**	**Organisation**
Chief investigator/s		
Research funder	This research is being funded by...	

What is the study about?
You are invited to participate in a study examining the current skills and training needs of the Community Care workforce in the region. We hope to learn more about your experience as direct care workers, the work that you and how your training/education has prepared you to do it.

Do I have to participate?
Being part of this study is voluntary. If you want to be part of the study we ask that you read the information below carefully and ask us any questions. You can read the information below and decide at the end if you do not want to participate. If you decide not to participate this will not affect your relationship with La Trobe University or any other listed organisation.

Who is being asked to participate?
You have been asked to participate because you are a direct care worker at one of the five participating community care service providers and we are interested to hear more about your experience of the role that you have.

What will I be asked to do?
If you want to take part in this study, we will ask you to complete a survey and if interested take part in an interview to further expand on the information provided through the survey. The survey consists of 50 multiple choice questions and should take approximately 30 -40 minutes to complete. If you decide to take part in an interview this will take approximately 30–60 minutes of your time. You will be asked to sign a consent form if you choose to participate in this part of the project. The interviews will take place at your workplace and your employer will allow time for you to take part.

What are the benefits?
The benefit of you taking part in this study is that your training and education needs will be identified to make sure that you have an opportunity to develop the ongoing skills and knowledge to meet the

demands of your role. The expected benefits to society in general are that the rural communities in the region will benefit from improved community care delivery and be provided with the best opportunity to improve health, wellbeing and independence.

What are the risks?
With any study there are (1) risks we know about, (2) risks we do not know about, and (3) risks we do not expect. If you experience something that you are not sure about, please contact us immediately so we can discuss the best way to manage your concerns.

Name/Organisation	Position	Phone	Email

What will happen to information about me?
We will collect and store information about you in ways that will not reveal who you are. This means you cannot be identified in any type of publication from this study.

We will keep your information for 5 years after the project is completed. After this time we will destroy all of your data. We will collect, store and destroy your data in accordance with La Trobe Universities Research Data Management Policy which can be viewed online using the following link.

The information you provide is personal information for the purposes of the Information Privacy Act 2000 (Vic). You have the right to access personal information held about you by the University, the right to request correction and amendment of it, and the right to make a compliant about a breach of the Information Protection Principles as contained in the Information Privacy Act.

Will I hear about the results of the study?
We will let you know about the results of the study by providing a report to your employer and by conducting an information session on completion of the study. If you would like a copy of the results you can request a copy by contacting Associate Professor Suzanne Hodgkin via e-mail, mail or telephone.

What if I change my mind?
You cannot withdraw consent once the **survey** has been returned as the survey responses are anonymous. You can choose to no longer be part of the interview component of the study at any time. You can request that your interview response be withdrawn and you can request that the interview is stopped without providing the researcher with an explanation. You can let us know by:

- Completing the 'Withdrawal of Consent Form'.
- Calling us.
- Emailing us.

Your decision to withdraw at any point will **not** affect your relationship with La Trobe University or any other organisation listed. Any identifiable information about you will be withdrawn from interview component of the research study. However, once the results have been analysed we can only withdraw information, such as your name and contact details. If results have not been analysed you can choose if we use those results or not. Once the results have been published the interview data cannot be withdrawn.

Who can I contact for questions or want more information?
If you would like to speak to us, please use the contact details below:

Name/Organisation	Position	Phone	Email

What if I have a complaint?
If you have a complaint about any part of this study, please contact:

Ethics Reference Number	Position	Phone	Email

Further information: https://www.latrobe.edu.au/researchers/research-office/ethics/human-ethics.

Annex 2

Comparison between Quantitative and Qualitative journal articles

Annex 2

COMPARISON BETWEEN QUANTITATIVE AND QUALITATIVE JOURNAL ARTICLES

	Quantitative	Qualitative
Title	Informative, attract the reader's attention, should accurately reflect the nature and focus of the study.	Informative, attract the reader's attention, should accurately reflect the nature and focus of the study.
Abstract	Short summary, provides an overview of what the research is about, what was done, how it was done, what was found, and what the results mean.	Short summary, provides an overview of what the research is about, what was done, how it was done, what was found, and what the results mean.
Keywords	6-8 keywords used to draw the reader's attention, also used to locate articles in electronic databases.	6-8 keywords used to draw the reader's attention, also used to locate articles in electronic databases.
Introduction	Brief overview of previous relevant research, provides a rationale for the study and an outline for what the research is aiming to do. Authors highlight a gap in knowledge, and describe what their study will provide in relation to this gap.	Brief overview of previous relevant research, rationale, provides a rationale for the study and an outline for what the research is aiming to do.. Authors highlight a gap in knowledge and describe what their study will provide in relation to this gap. In qualitative journal articles, the theory or framework used may be introduced.
Methods	Summarises the procedure, providing enough detail that another research study could replicate it including: participants, materials, study design, procedure and the process of data collection and analysis.	Discusses the design of the research and method employed. Explanation of why a qualitative approach has been used, including discussion of ontology and epistemology. Methods used to ensure rigour are discussed. Detailed description of research site, population, participants, and the process of data collection and analysis.
Results/Findings	Summarises the data collected and statistical analyses performed. Should report the results without any type of subjective interpretation. Some research intends only to describe the results for the sample, while other research attempts to make inferences about the population from the sample	Findings (results) can be presented as themes, verbatim quotations are used to elaborate on the explanation of the findings. Findings can be presented in different ways e.g. given without interpretation, or given with interpretations but a detailed discussion is left to the discussion section, or findings and discussion can be combined.
Discussion	Summarises and interprets findings, relates the findings back to previous research, considers the original research question or hypothesis, and discusses the clinical implications for the client and the profession.	Summarises and interprets findings, relates the findings back to previous research, literature, and theoretical framework. Considers the original research question or hypothesis, and discusses the clinical implications for the client and the profession.
Conclusion	Provides any limitations of the research and recommendations for future research	Provides any limitations of the research and recommendations for future research.

Glossary

Research Terms

action research

Usually conducted within an organization by people who conduct the research as part of the activity they are doing. They then analyse the findings to improve their practice – the way they do things.

aetiology (or etiology)

Refers to the cause of something or its origins.

anecdotal evidence

Evidence gathered (by chance) without being part of a research design. It's usually only from a small number of reports, so, even if it's correct, it can only be considered interesting rather than as substantial evidence of something. Anecdotal evidence can raise interesting questions which researchers will want to answer. Hence, quite often, it's the starting point of a properly designed study.

bias

A biased sample is one that is not representative; that is it does not represent the population to which the researcher is hoping to generalise. e.g. If the researcher wants to know how much sugar university students eat and only interviews members of the university basketball team, the results will be biased. The basketball players might be likely to eat more sugar than the average student. Another type of bias might be in the literature that the researcher refers to in the paper, only choosing studies that agree with a particular point of view.

bibliographical (or bibliography or reference list)

Details of books, journal articles or other sources which will help your reader to find them (so they can read them for themselves).

blinding

Subjects in a study are considered "blinded" if they do not know which group they are in. For example, in a study comparing the effects of two drugs, participants won't know whether they are taking drug A or drug B. A "double blind" study is when neither the subjects nor the researchers know who is taking which drug.

boolean operators

and/or/not. They help you to filter your literature search. For example, if you're searching for information on bone density in older people, you might search for bone density AND (older OR elderly OR geriatric) NOT arthritis.

case study

An in-depth investigation and analysis of something very specific.

citation

Mentioning or referring to sources (references) in a text. This can be direct or indirect quotation. APA and Harvard use name-date. Other systems use numbers.

cohort

A group of subjects with the same or similar characteristic (or experience). For example, research into the benefits of breastfeeding on sleep patterns in newborns might study a cohort of newborns. Another example might be a study of migraine sufferers where researchers compared one cohort who had never worn glasses with another cohort who had worn them.

control group

When an experiment is looking at the changes in a **dependent variable**, one group is exposed to the dependent variable and the control group is not. For example, researchers looking at the effects of independent gentle exercise in recurrent frozen shoulder sufferers might study one group who follows a prescribed exercise program and a control group that receives physiotherapy instead.

deductive reasoning

Arriving at a conclusion by working from a theory towards finding the evidence.

dependent variable

A variable that we consider as being an 'effect' is called a dependent variable (because the value of this variable depends on the "cause"). For example, consider the research question:

"Do carbohydrates aggravate gastritis? The researcher thinks that carbohydrates are the cause, and aggravated gastritis is the effect. The severity of the gastritis depends on the amount of carbohydrates.

Cause = carbohydrates (**independent variable**)

Effect = aggravated gastritis (dependent variable)

empirical

Can be seen through direct observation or experience rather than through theories or assumptions.(eg. Empirical evidence).

ethics

Being honest with participants and with other researchers about the purposes, intentions, risks and benefits of the research. This includes obtaining permission (consent) from participants to take part in a study and making sure they fully understand what will happen and how the results will be used. It also means treating participants well.

ethnographic

An approach which seeks to gather data from people in their normal environment. Usually, the researcher participates in some way in the data gathering process.

etiology

See **aetiology**

experimental

A method where the researcher controls or changes an **independent variable** to assess potential changes in a **dependent variable**. For example, does exercise (independent variable) cause a change in blood pressure (dependent variable)

focus group

A group of people selected to discuss their views or experiences regarding a specific issue or item. The conversation can generate ideas that might not come up in an individual interview.

grounded theory

Instead of starting with a theory and then trying to find evidence to support it, grounded theory is formed from the analysis of the data. The study may start with questions rather than a clear **hypothesis**.

hypothesis

A proposition put forward by the researcher which is evaluated by the collected data. What the research is trying to discover. It's usually a question framed as a statement. For example, for the question "Do school students eat more sugar than university students?" the hypothesis might be "School

students eat more sugar than university students". The research will then look for evidence to help support or not support this.

independent variable

A variable that we think is a 'cause' is known as an independent variable (because its value does not depend on any other **variables**). For example, consider the research question: Do carbohydrates aggravate gastritis? The researcher thinks that carbohydrates are the cause, and aggravated gastritis is the effect. The severity of the gastritis depends on the amount of carbohydrates. The independent variable is controlled by the researcher.

Cause = carbohydrates (independent variable)

Effect = aggravated gastritis (**dependent variable**)

inductive reasoning

Arriving at a conclusion or principle by generalising from specific facts or findings – starting from an observation and working up towards formulating a theory.

longitudinal

Study repeated with the same group of subjects over a long period of time.

meta-analysis

An analysis of the results of a large number of similar studies comparing and combining them to get a more complete picture.

methodology

The ideas that lie behind the design of the research. For example, a qualitative research methodology values holistic data collection with a lot of rich information, rather than numbers.

methods

How the data is collected e.g. a common research method, using a qualitative methodology is conducting interviews. Interviewing is a method.

n

The code used to refer to the number of subjects in a study. For example, if 67 nurses were interviewed to discover their views on communicating with bereaved relatives, any tables or statistical content about them in the study would say n=67.

narrative review

The use of stories as data.

null hypothesis

It's quite difficult to conclusively prove that something is true. Research is often designed to show the absence of evidence. If, for example, we want to know whether liquid analgesic is faster-acting than capsules, our null hypothesis would be that response time is the same. If our findings showed some difference, we could reject the null hypothesis. In other words, we found no evidence that they're the same, and our study shows that they might actually be different. The null hypothesis predicts no relationship or no difference between groups/treatment. Researchers accept or reject the null hypothesis based on the **p-value** as a result of statistical tests.

objective

A scientific, impersonal approach based on observation and measurement. The researcher's opinion is not included.

peer-reviewed

Reviewed and evaluated by other experts in the field who know the work.

phenomenology

In qualitative research, the study and understanding of human conscious experience.

phenomenon (one) phenomena (more than one)

A specific event or happening

P.I.C.O.

An abbreviation which helps to formulate clear research questions:

P = population; I = intervention; C = comparison; O = outcome

For example, Do asthmatic children (P) need more exercise (I) than non-asthmatic children (C) to enable them to breathe normally at night (O)?

placebo

A "fake" drug or treatment which appears identical to the real one. Usually used with **control groups** to compare against a group using the real one.

Primary research (data)

Information and /or data collected and /or observed directly from the research project. An original study.

P-value

Outcome of statistical significance testing. The probability that the same result may have occurred by chance. So, the lower the p-value, the more likely that the **hypothesis** is true. (Or, to put it another way, the lower the p-value, the less likely that the **null hypothesis** is true and therefore the null hypothesis is rejected).

qualitative research

An approach to research design that emphasises the non-numerical and uses more judgement and interpretation. This often uses extensive notes or interviews. An approach to research design which focuses on gathering data by talking to individuals, for example in interviews, **focus group**s and asking open-ended questions.

quantitative methods

An approach to research design that emphasises the collection of numerical data and the statistical analysis of hypotheses proposed by the researcher. An approach to research design which focuses on gathering data by counting or measuring, for example in **survey**s, questionnaires andasking yes/no questions. Another common quantitative method is taking measurements of participants e.g. measuring their height and blood presuure.

randomised

In an experiment, subjects are assigned to their groups by using a random assignment method. For example, tossing a coin is one procedure for assigning subjects randomly (without intention) to two groups.

randomised controlled trial

A type of study design in which there are usually two groups and participants are assigned to one of the groups, randomly. These groups are then called the **control group**, which does not receive the intervention and the intervention group which does. However, researchers do aim to keep the groups as similar as possible in terms of the participants' characteristics i.e. age, sex etc. (and often other characteristics that are important to the trial). The goal is to try to ensure that it is the intervention that is responsible for any observed effects after treatment and not some other factor.

reliability

The extent to which a test or measurement result is reproducible; that is, if the experiment were conducted again by the same or other researchers, the same or similar results could be expected.

robust

Can stand up to scrutiny. Measures what is says it measures.

scientific method

Objective systematic observation and measurement. The scientific method tries to ensure that personal views are excluded and that only scientifically gathered data and analysis form the basis of the study. Usually includes these stages: question; **hypothesis**; prediction; testing; analysis.

secondary research

Research into the **primary research** of others – not first-hand **experimental** work. Often a compilation of other people's work.

socially constructed

This refers to a belief or behaviour that has been learned by a person through interaction with others and common sets of ideas and practices. For example some people believe that gendered behaviours e.g. boys liking trucks are "socially constructed" rather than naturally occurring.

subjective

An approach incorporating personal feelings, views and judgements. Some disciplines such as counselling rely on these characteristics.

survey

A method which relies on questionnaires and/or interviews to collect data.

thick description

Describes not only a particular fact or behaviour but gives rich information about the context or framework in which it occurs. This might be through extensive notes or interviews. Thick description is associated with **qualitative research**.

triangulation

Using more than one method of data gathering to confirm findings.

truncation

You can use an asterisk (*) to replace a part of a word if you want to find variations when performing a literature search. For example, cathete* might give results for catheterize, catheterization, catheter, etc.

validity

The extent to which a test measures what it is intended to measure. E.g. If the researcher wants to know about levels of oral hygiene among school children, a useful question might be “How often do you brush your teeth?” This is valid because it contributes to the measure of oral hygiene activity. An invalid question might be “What colour is your toothbrush?” This might tell us whether or not the child owned a toothbrush, but nothing about whether and how effectively it was used.

variables

The characteristics or features of something which might vary and would need to be taken into account. For example, if the researcher is investigating the time taken to travel to university, variables might include distance from home, method of transport, weather or time of day.

wildcard

Replacing letters or characters with ? to get alternatives. For example, m?n, would give results for man and for men. Replacing whole words with ? can be useful. For example, you could look for sepsis ? if you wanted to find stuff on symptoms and treatment and prevention, etc., etc., (but you might get more results than you want!)

Some of these definitions are based on the following sources:

Boswell, C., & Cannon, S. (2014) Introduction to nursing research, Burlington MA, Jones & Bartlett Learning

Polar, S., & Thomas, S., Glossary of research terms, course reading online, La Trobe University Library