INTRODUCTION TO RESEARCH IN THE HEALTH SCIENCES

FIFTH EDITION

Stephen Polgar
BSc(Hons) MSc
Senior Lecturer, School of Public Health, Faculty of Health Sciences, La Trobe University, Melbourne, Australia

Shane A. Thomas
DipPubPol PhD MAPS
Director of Primary Care Research, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia

CHURCHILL
LIVINGSTONE
ELSEVIER

EDINBURGH LONDON NEW YORK OXFORD PHILADELPHIA ST LOUIS SYDNEY TORONTO 2008

CHURCHILL
LIVINGSTONE
ELSEVIER

First edition 1988
Second edition 1991
Third edition 1995
Fourth edition 2000
Fifth edition 2008

ISBN: 978-0-443-07429-5

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
A catalog record for this book is available from the Library of Congress

Notice
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our knowledge, changes in practice, treatment and drug therapy may become necessary or appropriate. Readers are advised to check the most current information provided (i) on procedures featured or (ii) by the manufacturer of each product to be administered, to verify the recommended dose or formula, the method and duration of administration, and contraindications. It is the responsibility of the practitioner, relying on their own experience and knowledge of the patient, to make diagnoses, to determine dosages and the best treatment for each individual patient, and to take all appropriate safety precautions. To the fullest extent of the law, neither the Publisher nor the Authors assume any liability for any injury and/or damage to persons or property arising out of or related to any use of the material contained in this book.

The Publisher

Transferred to Digital Print 2010

Printed and bound in the United Kingdom

Contents

Preface

Consistent with previous versions, the 5th edition of this textbook has two overall aims:

1. To introduce the fundamental principles of research methodology and explain how these principles are applied for conducting research in the health sciences.
2. To demonstrate how evidence produced through research is applied to solving problems in everyday health care.

It is widely recognized that the advancement of knowledge and practice in an area of health care depends on the education of individuals who are competent in undertaking health research and evaluations. We recognize that only a relatively small proportion of health sciences professionals have the interest or the opportunity to undertake full-blown research. At the same time, all contemporary health professionals participate in a 'culture' which privileges knowledge based on the critical review of scientific evidence. This textbook will enable students to participate in discussions concerning health problems and scientific approaches to treatment.

It must be emphasized that, in its 5th edition, *Introduction to Research in the Health Sciences* remains an introductory textbook. In order to retain the simplicity and readability of the text we have made the following changes:

- Inclusion of an additional chapter which focuses on the use of qualitative methods in health research.
- Inclusion of a new chapter for discussing systematic reviews and meta-analyses.
- More revision questions, in particular ones which emphasize the utilization of research for the solution of health problems.
- Revised and updated clinically relevant research designs such as single case studies and focus groups.
- Expanded discussion of effect size and clinical significance of quantitative data.
- Improvements in the presentation and formatting of the text, in particular by updating the figures and diagrams.

We would like to acknowledge our colleagues who have contributed in various ways to the development of this book since it was first published in 1988. We are grateful to our readers and students who have made helpful comments, making this book relevant for introducing students to health research. We are particularly grateful to Joanna Ng for her hard work and dedication in organizing the text for this 5th edition.

Melbourne 2007

Stephen Polgar
Shane A. Thomas

Section One

The scientific method

SECTION CONTENTS

In this section we will examine some of the basic characteristics of the scientific method. A 'method', in the present context, refers to a system for acquiring knowledge and establishing its truth. Health providers justify their theories and practices on the grounds that they are 'scientific', that is, based on scientific methodology. The scientific method is essential for conducting research and evaluation aimed at producing evidence, improving the effectiveness and cost-effectiveness of health services.

A common view of the scientific method is that it enables us to describe, predict, explain and perhaps to control events in the world. As an example, consider how researchers and practitioners in the health care system responded to the outbreak of the acquired immuno-deficiency syndrome (AIDS) epidemic in the early 1980s. One of the first signs of this epidemic was the clinical observation that young men presented with a deadly cancer (Kaposi's sarcoma) which was previously found only in the elderly. This was the initial evidence which indicated that further facts and explanations were required. It was hypothesized that the premature failure of the immune system was responsible for the disorder. Subsequent clinical evidence supported this hypothesis, leading to a clear definition of the disease as 'AIDS'. As more cases of AIDS were identified it became evident that some groups in the population were most at risk: persons who were involved in multiple homosexual relationships or who were intravenous drug users, or those who required blood products, for example, for haemophilia. An important development was the confirmation of the hypothesis that the disease was caused by a specific virus labelled 'HIV' (human immuno-deficiency virus). On the basis of the above clinical, epidemiological and virological evidence it was hypothesized that the HIV was transmitted through body fluids such as blood or semen. In this way it became possible to predict which practices in the population involved a high level of risk for transmitting the virus.

The above findings and hypotheses contributed to constructing theories of AIDS, providing systematic and integrated conceptual frameworks for explaining the transmission and the clinical features of the disorder. These theories also inform current practices aimed at controlling the epidemic, such as testing blood products for HIV, or promoting 'safe sex' behaviours. The effectiveness of these practices (at least in some countries) for containing the epidemic provided evidence confirming the accuracy

of our understanding of the problem. At the same time, we remain sceptical about certain aspects of our theories, recognizing that there are as yet no vaccines or pharmacological cures for AIDS.

You might have recognized that the above brief account of attempting to contain AIDS is deficient in an essential way: it excludes evidence concerning the personal experiences and actions of persons at risk of or having HIV. The above account represented a quantitative approach to the scientific method aiming to provide a mechanistic or reductionistic explanation of health and illness. This method is not appropriate for researching and theorizing how persons, their families and the community respond to health-related issues and problems. Research methods should also include qualitative or interpretive approaches to discover the personal meanings involved in being at risk of or suffering from diseases or disabilities. Qualitative methods are also appropriate for researching the cultural context and social construction of a disorder; for instance, discovering the images of AIDS sufferers communicated by the media in a given community.

The position taken in the present book is that the scientific method, as applied to health care, must include both quantitative and qualitative methods. These methods are used to produce evidence for solving health-related problems. Regardless of controversies regarding the nature of science and its methodology, there is a broad consensus concerning the principles and rules for producing scientific evidence. Our goal is to convey these principles and rules to our readers.

Chapter One

1

Methods and health research

CHAPTER CONTENTS

Introduction

Health research is a systematic and principled way of obtaining evidence (data, information) for solving health care problems and investigating health issues. Research is *systematic* in that researchers tend to follow a sequential process (Fig. 1.1) and *principled* in that research is generally carried out according to explicit rules. These rules or principles constitute the method.

The primary aim of this chapter is to outline the key characteristics of methods for conducting enquiry. Special emphasis is placed on method as a means for conducting applied research and producing the best possible evidence for solving health problems.

The general aims of this chapter are to:

1. Define what is meant by a method, and outline some common methods of enquiry.
2. Draw attention to some controversies concerning the nature of scientific method.
3. Compare qualitative and quantitative approaches to research.
4. Discuss how method is applied to conducting health sciences research.

Methods and knowledge

Patient care involves the acquisition of a set of specific skills, the practice of which is justified in

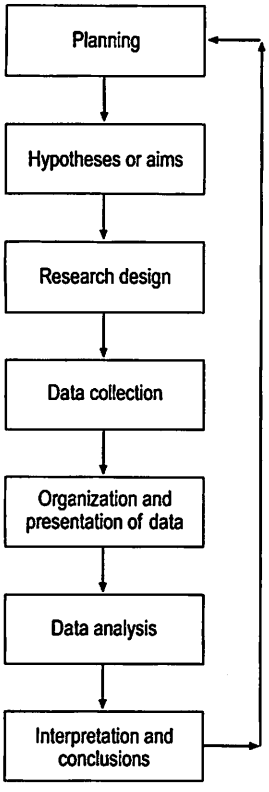


Figure 1.1 • The research process.

terms of a body of professional knowledge. This body of knowledge is based on the use of appropriate methods. In general the term 'method' refers to a systematic procedure for carrying out an activity and in the present context implies a set of rules which specify:

1. How knowledge should be acquired.
2. The form in which knowledge should be stated.
3. How the 'truth' or validity of the knowledge should be established.

Before we begin a discussion of the scientific method it is useful, as a means of contrast, to look at some of the other methodological approaches.

Authority

According to this method, knowledge is considered true because of tradition, or because an experienced and distinguished clinician says that it is true. As a student, you are often asked to accept statements as true because your teachers and clinical supervisors say they are true. To maintain their authority, the 'sources' of knowledge acquire and cultivate various signs of expertise, such as the appellation of 'Professor', or encourage the performance of status rituals. For example, consider the 'Consultant' who sweeps into a hospital ward followed by a retinue of students, registrars and nurses. Who would dare to question the truth of any of the consultant's sacred utterances? However, there are problems with the method of authority: what happens when the statements arising from one authority are contradicted by those made by an equally prestigious authority? For example, say Authority X claims that in their experience psychoanalysis is effective, while Authority Y states that the technique is useless. How can we resolve such conflicting claims? In practice, unless objective and acceptable criteria for resolution can be found, we will have unending arguments, *ad hominem* (personal) attacks, or in instances of some religious or political disagreements, violence. That is, the controversy is resolved by denigrating or silencing the dissenting authority. In contrast, the scientific method emphasizes the evaluation of the evidence for establishing the truth of conflicting statements.

Reasoning

Reasoning is commonly used to arrive at true knowledge. It is assumed that if the rules of logic are applied correctly, then the conclusions are guaranteed to be valid. As an example, let us look at the following syllogism:

1. All persons suffering from heart disease are males.
2. Person X has heart disease.
3. Therefore, person X is a male.

Logic guarantees that the conclusion (3) is true, provided that the syllogism is in a valid form and the premises (1) and (2) are true. Clearly, the limitation of formal (that is, 'content-independent') reasoning is that it works in practice only if we have means for establishing the factual truth of the premises. In the above example, conclusion

(3) might be empirically false, given that the premise (1) is factually false.

Logic and mathematics are very much a part of science but we require strong evidence to support the effective logic and mathematical operations.

Intuition

Knowledge is sometimes acquired by sudden insights which arise without conscious reasoning. Truth is judged by the clarity of the experience and its emotional content (the 'Eureka!' experience). For example, after working with a patient without success, you might have a sudden insight about how to change your treatment programme. Sometimes these insights lead to new theories and treatments. Unfortunately, even the strongest intuitions are sometimes proven false by cold, boring facts. Newton has referred to the disappointment that occurs when a beautiful hypothesis is destroyed by an ugly fact.

Discoveries are sometimes resisted because they seem counterintuitive. Ignaz Semmelweis, a perceptive and humane 19th-century physician, noticed appalling levels of puerperal (child birth-related) fever and maternal death at his hospital. He argued that the infection was spread by medical students and staff who went to delivery rooms from the morgue without adequately washing their hands. In 1848, Semmelweis introduced antiseptic procedures in his wards and demonstrated a substantial reduction in mortality from puerperal fever. However, his colleagues were offended by the notion that physicians were carriers of disease. Semmelweis was dismissed from his post, ostracized by the medical establishment and died in pitiful circumstances. Clearly, in the light of contemporary knowledge we can say that Semmelweis' insights were correct, while his colleagues' intuitions were mistaken. Authority, logic and intuition all have their places in health care and research. In a general way the scientific method can be contrasted with other methods in that it emphasizes the need for evidence. What constitutes scientific evidence and what the evidence indicates are complex matters that are being addressed in this book.

The scientific method and positivism

The scientific method has crystallized over a period of several centuries, concomitantly with the growth of scientific research. The beginnings of modern Western science are generally traced to the 16th century, a time in which Europe experienced profound social changes and a resurgence of great artists, thinkers and philosophers. Gradually, scholars' interests shifted from theology and armchair speculation to systematically describing, explaining and attempting to control natural phenomena. These changed circumstances allowed philosophers such as Descartes and Francis Bacon to challenge tenets of medieval thinking, and scientists such as Galileo, Newton and Harvey to propose new models of natural phenomena.

The approaches of such great thinkers had three basic elements, which are the basis of scientific method:

1. *Empiricism.* The notion that enquiry ought to be conducted through observation and knowledge verified through evidence.
2. *Determinism.* The notion that events in the world occur according to regular laws and causes. The aims of research are to discover these rules and causes.
3. *Scepticism.* The notion that any proposition or statement, even those by great authorities, is open to analysis and critique.

The scientific method is represented in a simple form in Figure 1.2. This representation is consistent with a view of natural science methodology called 'positivism' by the 19th-century philosopher Comte, as well as 'reductionism'. Both these terms have acquired multiple meanings in the area of philosophy of science. Our interpretation of Figure 1.2 is explained below.

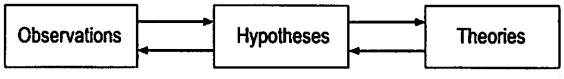


Figure 1.2 • The scientific method.

Observations, description and measurement

Considering Figure 1.2, let us start with observations. The description of phenomena involving the precise, unbiased recording of observations of aspects of persons, objects and events forms the empirical basis of all branches of science. Observations can be expressed as either verbal descriptions or sets of measurements (see Section 4). The perceptions of the investigator must be transformed into descriptive statements and measurements that can be understood and replicated by other investigators. Some research is based on observation made with instruments (such as recording electrodes, microscopes and standardized clinical tests), while other research calls for observation unaided by instruments. Although advances in instrumentation have contributed enormously to scientific knowledge, the use of complex instrumentation is not a necessary feature of scientific observation. Rather, the key attributes of scientific observation are accuracy and replicability by other scientists. When observations are appropriately summarized and are confirmed by others, they form the factual bases of scientific knowledge.

Generalization and induction

Statements representing observations or measurements are integrated into explanatory systems called hypotheses and theories. The logic underlying scientific generalizations is called induction. Induction involves asserting general propositions (hypotheses, theories) about a class of phenomena on the basis of a limited number of observations of select elements. For example, having observed that penicillin is useful for curing pneumonia in a limited set of patients, we make the generalization 'The administration of penicillin cures pneumonia (in all patients)'.

Hypotheses

The statement 'The administration of penicillin cures pneumonia' is an example of a hypothesis. Scientific hypotheses are statements that specify the expected nature of the relationship between two or more sets of variables. In this instance, the first variable relates to the administration of penicillin and the second variable is related to changes in clinical observations or measurements concerning pneumonia. As we shall see in subsequent sections, an important feature of scientific hypotheses is that the terms used must have clear-cut, observable referents. When these hypotheses acquire strong empirical support, they may be called laws.

Theories

Scientific theories are essentially conjectures representing our current state of knowledge about the world. Hypotheses are integrated into more general explanatory systems called theories. A theory will clarify the relationships between diverse classes of observations and hypotheses. For example, a theory to explain why drugs called antibiotics are effective in curing some infectious diseases integrates evidence from diverse sources such as microbiology, pharmacology, cell physiology and clinical medicine. Other examples of theories are the heliocentric theory of the solar system, the DNA theory of genetic inheritance, and the neuronal theory of central nervous system functioning. It is an essential feature of scientific theories that they are statements based on the correct use of language and logic. Some theories entail a model (see Fig. D1.1, p. 16), which is a mathematical or physical representation of how the theory works. In this way, theories specify the causes of events and provide conceptual means for predicting and influencing these events. In health care, theories are important for explaining the causes of health and illness and predicting the probable effectiveness of treatment outcomes.

Deduction

A scientific theory should lead to a set of empirically verifiable statements or hypotheses. In addition to being generalizations based on evidence, hypotheses are also deduced logically from the statements and/or mathematical models which specify the causal relationships postulated by the theories. For example, if we hold the theory that the patterns of activity of a set of neurones in the

occipital lobe mediate visual sensation in humans, then the hypothesis follows that the activation of these neurones (say, by electrical stimulation) will lead to the report of visual sensations. Such hypotheses have been the bases for subsequent spectacular clinical advances such as artificial vision through cortically implanted electrodes.

Controlled observation

It is desirable that hypotheses are tested under controlled conditions. The aim of control is to discount other competing hypotheses for explaining the predicted phenomenon. For example, if we intend to show that occipital lobe stimulation causes the visual sensations, we must show that we are controlling for any type of brain stimulation causing such changes. Conversely, we would need to show that occipital lobe stimulation does not lead to a host of other types of sensations. Only by discounting alternative explanations through control can we have confidence in the relevance of our observations for our research hypothesis. Of course control must be ethical, as we shall see in later chapters.

Verification and falsification

An essential characteristic of a scientific proposition is that it should be 'falsifiable'. That is, there should be a clear empirical outcome that could, if found, show that the proposition was false. For example, consider Einstein's hypothesis that $E = mc^2$. If laboratory researchers found that the energy locked up in matter corresponded to, say, $E = mc^3$, then Einstein's hypothesis would be shown to be false. This seems not to have happened yet, so that $E = mc^2$ is falsifiable but not false. After the evidence has been collected, the investigator decides whether or not the findings are consistent with the predictions of the hypothesis. If the hypothesis is supported by the evidence, then the theory from which the hypothesis was deduced is strengthened or verified. However, when the data do not support the hypothesis, the related theory is falsified. If a theory can no longer predict or explain evidence in its empirical domain, it becomes less useful and is usually later discarded in favour of new, more powerful theories. Therefore, scientific theories are not held to be absolute truths, but rather as provisional explanations of available evidence.

The application of the above process has contributed to the spectacular growth of scientific knowledge. Observation and measurement, facilitated by new instrumentation, resulted in the discovery of an enormous number of accurate and reproducible facts about health and illness. New facts both challenge existing theories and call for the creation of novel, more powerful theories. The new theories serve as impetus for more research, resulting in new instrumentation and observations. Advances in scientific knowledge have been applied to creating new technologies, which in turn contribute to new discoveries and advances in scientific knowledge. For example, the invention of computers was possible because of advances in electronics, chemistry and mathematics. In turn, the use of computers is now contributing to making and summarizing scientific observations or formulating explanatory models. In addition, the use of computers as information-processing systems has generated useful metaphors for theoretical advances, such as explaining the human brain and mental functioning. In this way, the scientific method contributes to advancing theory and practice, helping us to describe, explain, predict and sometimes control the world in which we live.

Controversies concerning the nature of method: post-positivism

The above description is not the only interpretation of scientific method. Rather, there are good reasons why the positivist view of the scientific method was challenged in the second half of the 20th century.

First, the scientific method is a set of rules devised and applied by philosophers and scientists. They are not eternal truths, but conventions believed to be useful for conducting scientific enquiry. In this way, not only the content but also the methodology of science is open to criticism, debate and change.

Second, the interpretation of what constitutes the scientific method is an activity pursued by philosophers of science and epistemologists. In exploring different conceptual frameworks (realism, instrumentalism, anarchism, idealism, etc.) concerning the nature of reality and knowledge, philosophers generate unending controversies over the nature of the scientific method.

Third, advances in science contributed to changes to our methods and the ways in which we explain the world. For example 'quantum theory' and the 'theory of relativity' challenged the mechanistic, clockwork-like view of reality. Modern theories of physical reality are fuzzier and more probabilistic and the observers are not seen as being outside the phenomena they are studying.

The following ideas represent key aspects of the post-positivist view of scientific methodology.

The theory-dependence of observation

Critics such as Chalmers (1976) argued that it is simplistic to believe that observations are made independently of theoretical notions held by the observer. The observer is selective with regard to what is recorded as evidence. Our observations and facts are 'theory-dosed', that is, theories specify what observations are of importance and what aspects of these observations should be recorded or ignored. Schatzman & Strauss (1973) put this point elegantly when they stated that the researcher 'harbours, wittingly or not, many expectations, conjectures and hypotheses which provide him with thought and directives on what to look for and what to ask about'. For instance, in observing the electroencephalogram (EEG) of an epileptic patient, our perception is guided by theories of the electrical activity of the brain and the nature of brain pathology. We will also hold ideas about how the EEG machine works (for example, electrode sensitivity, amplification, etc.) and identify artefacts in the evidence. What is observed as evidence of epilepsy by an expert could be perceived as meaningless squiggles by a naive observer. We shall take account of this point later in this book, when we discuss the reliability and validity of measurement.

The validity of induction

Philosophers of science have questioned the logical validity of making general claims on the basis of a limited set of observations. To quote Chalmers (1976): 'any observational evidence will consist of a finite number of observation statements, whereas a universal statement makes claims about an infinite number of possible situations'.

For example, we might have observed that the administration of penicillin cured pneumonia in 100000 patients. This does not necessarily guarantee the logical 'truth' of the universal statement 'penicillin cures pneumonia', or that patient number 100001 will be cured. We will look at the theories-related issue of generalization from samples in this book. Scientific theories are seen as probabilistic, in the sense that new, inconsistent evidence might emerge in the future, challenging the generality of the theory. Also, it has been argued by some philosophers of science that hypotheses *need not* be based on induction, but may arise from any source, provided that they have falsifiable empirical consequences.

What constitutes falsification?

It was stated earlier that when novel empirical evidence is inconsistent with the predictions of a theory, the theory is 'falsified' and is eventually modified or discarded. Commentators such as Lakatos (1970) argued that theories are not, in practice, so readily modified or discarded by scientists. Rather, they are structures which have an inner hard core of propositions protected by an outer belt of auxiliary, modifiable hypotheses. Evidence inconsistent with predictions based on the theory results in the modification of auxiliary hypotheses, rather than discarding the 'core'. Consider, for example, the 'germ' theory of infectious disease, on the basis of which one would predict that penicillin (which kills germs) will cure bacterial infections. Suppose that we administer penicillin to a number of patients with an infectious disease and find, contrary to what was predicted, no clinically useful changes. On the basis of this falsification, will we discard the germ theory of disease? No. Rather, we will utilize an auxiliary hypothesis to explain our findings, such

as 'the development of penicillin-resistant bacteria'. The methodological issue which remains controversial is the logical basis for discarding one theory and accepting its rival (see Feyerabend 1975). As we shall see, we judge the outcome of a research programme in the context of an overall pattern of related findings and theories.

Science and the cultural context

Science as a human activity

An important aspect of post-positivism has been the recognition that human values are an integral part of scientific inquiry. Scientific enquiry is conducted in particular social settings, by individuals with personal aims and values. Some more recent formulations of methodology take into account the social and interpersonal conditions which influence scientists' professional activities (Feyerabend 1975, Kuhn 1970). In this text, we will pay attention to social values, in the context of ethical considerations for designing and conducting research (Ch. 2). Also, we recognize that the formulation of hypotheses and theories are creative acts, rather than the outcomes of the automatic application of induction. In this way, the questions asked by health researchers and the ways in which they explain their findings are influenced in subtle ways by the cultures in which they live and work.

To understand the nature of scientific research in general, and health research in particular, it is useful to examine how they fit into an overall social context. That is, the way in which persons living in a society view health and illness and the ways in which health workers carry out their professional roles influence the range and scope of health care research (e.g. Taylor 1979). Until recently, the 'medical model' was by far the most dominant approach to understanding illness in Western society. Briefly, in terms of the medical model, illness is represented by a particular lesion or dysfunction within the human body. The role of the health professional is to identify the location and nature of the lesion or the clinical imbalance and to implement appropriate measures to correct the problem. The patient is assigned a rather passive role in this process, being the 'locus' of the lesion or imbalance and a person complying with the health professionals' recommendations. In the context of the medical model, the most appropriate research is seen as that which improves the technical effectiveness and, therefore, the social power of the health professional.

In Western society the medical model has been, and will continue to be, an influential model for guiding clinical practice and health research. Nevertheless, there have been gradual changes in health care that require the questioning of the generality of the medical model for the following reasons:

1. There have been important changes in the roles and status of health professionals such as nurses, occupational therapists, physiotherapists, speech pathologists and podiatrists. Rather than taking an ancillary, paramedical position, these individuals are becoming directly and independently responsible for a broad range of health care functions, including prevention, assessment, therapy and rehabilitation. However, the perspectives and practices of these health professionals are at times quite different from those involved in the medical model and may require quite different approaches to research and theory. The gradual establishment of university-based education for these professionals has provided an increased opportunity for relevant research, which is not necessarily guided by the medical model.
2. From the 1970s onwards it had become evident that there had been a significant increase in the cost of health care, in part due to the adoption of very expensive diagnostic procedures and the increasing trend for specialization among medical practitioners (Taylor 1979). Attempts to contain health care costs have resulted in several lines of approach, including:
   (a) increased focus on preventive health care, as shown by anti-smoking campaigns, or drink driving legislation to contain the incidence of road trauma
   (b) increased efforts to enhance the cost-effectiveness of current treatment strategies; for example, by improving communications between clinicians and patients or managing patients' fears concerning treatments through the application of psychological techniques
   (c) increased support for self-help groups and community-managed health centres with

a more active involvement of individuals in understanding and managing their own health.

3. Preventive and educational approaches require somewhat different views of the client and the professional to that implicit in the medical model. In turn, research is required which can clarify the relationships between community lifestyles, individual behaviours, health and illness and general economic and social circumstances (e.g. Gardner 1989).

The more holistic approach, which also informs health care research, is called the *psychosocial* or *biopsychosocial* model (Engel 1977), and currently has had considerable influence in how we conceptualize health care and how we plan and carry out research. At the same time, the medical model strongly based on positivist thinking remains influential in research aimed at understanding the workings of the human body and improving the technical aspects of health care.

Pragmatism: the combined use of quantitative and qualitative methods

Not all scholars agreed with the notion that qualitative and quantitative methods could be logically combined. The controversy was resolved on pragmatic grounds; both methods are needed to solve applied problems in areas such as health care (McGartland & Polgar 1994).

Pragmatism is a system of philosophy that avoids speculation about nebulous abstractions such as 'The Truth'; rather it defines truth as what works (Tashakkori & Teddlie 1998). A pragmatic, problem-solving approach is followed in the present book, emphasizing the need for both qualitative and quantitative methods for effectively solving health problems.

The introduction of a biopsychosocial approach has raised questions about the combination of methodologies relevant to health sciences research. We perceive patients or clients in two different but interrelated frameworks: first, as broken down or malfunctioning biological systems; and second, as persons, like ourselves, living in a society and who are attempting to make sense of and cope with their particular health care problems.

As argued earlier, the first view informs a reductionist or *quantitative* approach to research and knowledge. That is, we view our patients objectively, as natural objects, and attempt to identify and measure important variables which represent the causes and expressions of a clinical condition. We develop models and theoretical frameworks which systematically explain how these variables are interrelated and undertake therapeutic actions which serve to diminish the variables representing illness or disability. Our therapeutic actions are the technical applications for our scientific theories; their outcomes and effectiveness should be tested under controlled conditions.

The second view informs a *qualitative* or interpretive approach to research and knowledge. We view our patients as persons and attempt to gain insights into their subjective experiences and the reasons for their actions in particular situations. We develop theories for interpreting the nature and development of their personal points of view, and to inform our therapeutic actions so that they seem meaningful and appropriate to our patients.

Take, as an illustration of the above approaches, a patient with cancer. In a quantitative analysis we attempt to quantify the problem by using appropriate instrumentation which measures variables such as the size and location of the tumour and the extent of its spread within the organism. Consistent with current theories of the nature of cancer, various techniques such as surgery, radiology or chemotherapy might be brought into play, the effectiveness of which will be judged in terms of controlling specific variables associated with the condition, such as levels of pain, weight or time of survival of the patient.

With a qualitative perspective, we may address the meaning of the condition from the patient's personal point of view, within the context of his or her family setting and social circumstances. That is, the patient's value systems must be clarified and understood before actions such as assessments and therapeutic actions are undertaken. Enquiry might uncover conflicting values, for instance, concerning the implementation of a programme of chemotherapy. From the clinician's point of view,

a radical programme of chemotherapy might seem appropriate, as probably quantitatively extending the patient's life by several years. However, from the patient's point of view, the quality of life under chemotherapy might seem inadequate, and also they might not wish to be a burden on their families. Clearly, both in clinical practice and in related research, evidence from both quantitative and qualitative enquiries should be integrated to ensure effective health care.

Although there are differences in how health scientists approach problems, Figure 1.1 shows the sequence of procedures commonly involved in quantitative research. You will find that the organization of this book follows the sequential stages of the research process, as outlined below.

1. *Research problems and questions.* The first step in problem solving is to state, clearly and unambiguously, the problem which we are attempting to solve. In general, research problems reflect both the unmet health care needs of patients and opportunities created by new theoretical and technological advancements for improving health care. Having identified a problem (or opportunity) our next step is to ask the 'right' question. A well-formulated research question will guide the research project for producing the evidence to answer the question and solve the research problem.
2. *Planning.* As we shall show in Section 2, research planning involves selecting appropriate strategies and measurements to answer questions or to test hypotheses. It will be argued in Chapter 2 that planning must take into consideration previous research evidence as well as ethical and economic factors before the appropriate research strategy and measurement are selected, and the precise research hypothesis or aim is stated.
3. *Research design.* The usefulness of research depends on plans specifying appropriate sampling methods for ensuring the generalizability of the results. Appropriate designs are selected to ensure controlled observation in order to demonstrate causal relationships.
4. *Data collection.* The next step in the research process is collection of data. We will examine data collection methods commonly employed in qualitative and quantitative health research and the appropriate ways of carrying out measurements, and different types of measurement scales available for research and clinical assessment.
5. *Organization and presentation of the data.* Section 5 introduces descriptive statistics representing the conventions for summarizing and describing the data. Chapters 13–16 examine basic concepts in this area, outlining how graphs and various descriptive statistics are used to condense and communicate research and clinical findings. The presentation and analysis of qualitative data will be discussed in this text.
6. *Data analysis.* Data analysis involves applying the principles of probability for calculating confidence intervals and testing hypotheses. The area of inferential statistics, involving decisions concerning whether the data support experimental hypotheses, is outlined in this text.
7. *Interpretation of the evidence.* The final step in any research project involves interpreting one's findings. The findings may support existing theories or practices, suggest that new techniques may be more effective, or suggest new theoretical notions that are more able to explain phenomena. It is rare that the findings from any single research project are completely definitive, and often the results may suggest the need for further investigation in related subject areas or contexts. This issue will be examined in Sections 6 and 7.
8. *Publication.* For research to be scientifically meaningful, investigators must present their results in professional journals and at conferences. Research findings become part of scientific knowledge only if they stand up to methodological critique and replication. We will discuss conventions for preparing publications and outline steps for critically evaluating published research. Finally, in this text we will see how evidence from various related publications can be synthesized for guiding evidence-based practices.

Research and clinical practice

Research methods cover a wide variety of skills and techniques aimed at the methodologically valid investigation of questions of interest to the researcher. These methods of enquiry are not restricted to research laboratories, nor need they involve expensive equipment or large research teams. Rather, these methods imply an approach to stating and answering questions in any setting.

Applied research in the health sciences focuses on issues such as the prevalence and causes of illness, the usefulness and accuracy of assessment

techniques, or the effectiveness of treatments. Applied research which is published aims at producing findings that are of general interest to groups of professionals working in the health field.

Research methods interact with health practices in multiple and mutually productive ways. An important general aim of this textbook is to discuss the relationships between theories, practices and the ways in which research methods contribute to improvments in health care. The term 'evidence-based medicine' has been defined by Sackett et al (2000) as '... the integration of best research evidence with clinical expertise and patient values'. Evidence-based health care (EBHC) is a contemporary movement aimed at ensuring that health services are based on the best available scientific and clinical evidence. The fundamental question which is explored in this book is 'what constitutes best research evidence?' The evidence is, of course, produced by applied health research conducted in accordance with the principles of the scientific method. Clearly if you expect to participate in the programme of EBHC you will need to understand research methods sufficiently for making informed and critical judgments concerning the quality of the available evidence.

Finally, we need to keep in mind that the research is produced by the hard work of men and women who were undergraduate students, just like you. Ultimately it will be up to you (as nurses, podiatrists, physiotherapists, speech pathologists and occupational therapists) to carry out the research which underpins the effectiveness, advancement and prestige of your profession. If not you, then who?

Summary

There are several common methods used for acquiring, stating and establishing knowledge. The scientific method is one of these and it underpins the validity of diagnoses and clinical interventions in Western health care systems. The scientific method is concerned with applying a set of rules or conventions that will allow us to produce scientifically valid knowledge. These rules specify how observations should be made, and how theories and hypotheses should be stated and evaluated.

Theories and hypotheses obtained and verified through scientific enquiry are not held to be absolutely 'true'. An inherent part of the scientific approach is scepticism regarding both the contents of knowledge and the underlying methodology. We have pointed out that there are controversies concerning what constitutes scientific methodology. It was argued that the scientific method is directly applicable to conducting research in the health sciences. In general, the stages for research include planning, stating aims or hypotheses, and formulating designs; collecting, summarizing and analysing the data; and drawing conclusions. This process can be applied to ensuring advances in health care and to problem solving in specific clinical settings. EBHC is a current approach to using best available research evidence for delivering health care. There is a close relationship between professional practices and health sciences research. The rest of the book is organized to follow the stages involved in performing and applying health research.

Self-assessment

Explain the meaning of the following terms:

authority
data
deduction
determinism
empiricism
evidence-based health care
falsification
hypothesis
induction
intuition
observation
positivism
post-positivism
pragmatism
qualitative
quantitative
rationalism
research programmes
scepticism
theory

True or false

1. A 'method' specifies how knowledge should be acquired, stated and tested.
2. Very strong intuitions always turn out to be true.
3. The method of authority depends on the status and credibility of the source.
4. When a syllogism is in a valid form, the conclusions should be factually true provided the premises are true.
5. Rationalism calls for correct reasoning for acquiring truth.
6. The case of Ignaz Semmelweis illustrates that medical decisions are based on the scientific method.
7. Scepticism refers to the notion that all knowledge is false.
8. Empiricism refers to the notion that events in the world occur according to regular laws and causes.
9. Scientific observations are different from ordinary observations because they depend on the use of instruments.
10. Scientific observations must be recorded as numbers.
11. Hypotheses are unproven theories.
12. Hypotheses are statements which specify the nature of the relationship between two or more sets of observations.
13. The logic underlying scientific generalization is called induction.
14. Scientific theories represent notions of natural events and their causes.
15. A good scientific theory is one that, in principle, cannot be falsified.
16. Controlled observation aims to identify the causes of events.
17. Scientific theories should enable the deduction of empirically verifiable hypotheses.
18. When the empirical evidence is found to be consistent with the implications of a hypothesis, we can say that the theory from which the hypothesis was deduced is absolutely true.
19. When we say that a theory was falsified, we mean that hypotheses deduced from it were not supported by empirical evidence.
20. We can question the content but not the methodology of science.

Multiple choice

1. A scientific theory is a set of statements:
   a conforming to the rules of logic
   b explaining the relationships which pertain among apparently diverse phenomena
   c which lead to empirically testable hypotheses
   d all of the above.
2. The statement 'Persons who are highly anxious do not perform well on learning tasks' is:
   a a theory
   b a hypothesis
   c false
   d in principle empirically untestable.
3. A scientific hypothesis:
   a should be verified through the use of logic and disputation
   b should be open to empirical verification
   c cannot arise through intuition
   d *a* and *c*.
4. The results of scientific research:
   a should be made available for critique and replication
   b should not be used to support existing theories
   c must be obtained under controlled laboratory settings
   d must conform to public expectations about the outcome.
5. Descriptive statistics:
   a are based on the principles of probability
   b represent conventions for planning research
   c represent conventions for summarizing and organizing data
   d specify the selection of appropriate measurement scales.
6. The scientific method is a set of rules specifying how:
   a scientific knowledge should be acquired, stated and tested
   b scientists should conduct their life
   c society should conduct its affairs
   d all of the above.
7. Authority, as a method, is:
   a no longer relevant to the conduct of health care
   b the fundamental source of the scientific method
   c neither *a* nor *b*
   d both *a* and *b*.

8. A therapist without formal medical qualifications treated cancer patients using an 'alternative' regimen of herbs, massage and meditation. Say that there is no scientific evidence available that such a treatment is effective. It follows that:
   a the treatment was ineffective because it lacks scientific evidence
   b the treatment was ineffective because the practitioner was unqualified
   c both *a* and *b*
   d neither *a* nor *b*.
9. As evidence for the effectiveness of the treatment, the therapist provides 100 signed statements from current or former patients claiming that they were satisfied with the treatment. One reason such evidence lacks scientific validity is that:
   a the opinions of patients cannot, in principle, constitute scientific evidence
   b it was not acquired in a manner consistent with the principles of scientific observation
   c we cannot, in principle, generalize from only 100 observations
   d it pertains to an 'alternative', non-scientific treatment regimen.
10. In the context of the scientific method, evidence which would be most indicative of the effectiveness of the treatment would be:
   a the support of medically qualified people
   b the argument that each patient has the right to select their own form of treatment
   c that the survival times of the patients were better than those of equivalent patients being treated with conventional therapies
   d that the patients were willing to pay money to receive the treatment.
11. The statement 'If my theory of schizophrenia is true, then the sun will rise tomorrow morning' is:
   a probably based on the invalid use of deduction
   b untestable
   c based on the method of authority
   d *a* and *b*.
12. According to an astrologer, people who were born under a particular star sign are 'basically kind and very intelligent, although because of their modesty, not sufficiently appreciated by others'. To test the truth of this statement, the astrologer asks a group of individuals if this description fits their personality; 95% of the sample agrees that the description is accurate. One of the problems with this enquiry is that:
   a astrology is inherently false, therefore the evidence must be wrong
   b 'personality' is inherently a misunderstood concept
   c in this case a 100% agreement is required for acceptable evidence
   d it is contrary to the principles of controlled observation.
13. The notion that facts are 'theory-dependent' implies that:
   a what is recorded as evidence is independent of the event being studied
   b theories and hypothesis have no relationship to empirical observations
   c what is recorded as evidence can be influenced by the theoretical notions of the observer
   d only our minds exist, rather than a material world.
14. Lakatos argued that:
   a theories are discarded only when their 'hard core' of propositions are falsified
   b a 'belt' of auxiliary hypotheses protect the fundamental 'hard core' propositions of a theory
   c theories cannot be refuted or falsified on the basis of empirical evidence
   d both *a* and *b*.
15. Which is the *least* controversial statement concerning the scientific method?
   a There is only one acceptable form of the scientific method independent of the phenomenon being investigated.
   b Scientific theories are derived from the use of induction, rather than creative insights.
   c The principles of the scientific method can be usefully applied to conducting health research.
   d Scientific enquiry is conducted independently of the personal values and social environment of scientists.
16. Which of the following problems might be best approached through the scientific method?
   a The clinical effectiveness of a new instrument needs to be evaluated.
   b The hospital budget is cut by 10%.
   c Personnel refuse to staff an abortion clinic on moral grounds.
   d The nursing staff go on strike.

Section One

Discussion, questions and answers

These questions ask you to examine issues concerning the applications of the scientific method to various aspects of health sciences research. Unlike the multiple choice and true or false questions, these discussion questions do not necessarily have a single correct answer. Rather, they are aimed at promoting a critical, integrative view of conducting research.

Our first discussion question involves theories and examines the relationships between theories, models, hypotheses and empirical observations. Theories are conceptual frameworks (as we discussed in Ch. 1) integrating a range of related observations and explanatory hypotheses. We may deduce empirically testable hypotheses from our theories.

On the basis of observations, in particular when the observations are carried out under controlled conditions, we establish the probable truth of our hypotheses, and thereby support or falsify the theories which were originally the sources for the hypotheses.

Some theories include models which represent specific aspects of a theory. Models are used to explain real situations and predict novel empirical outcomes. Models can be as follows:

1. *Physical models*. These models are constructed from materials, for example a 'pump' model of the heart for showing how the circulatory system works, or a construction of the DNA molecule to show how different nucleotides are organized in order to replicate genetic information.
2. *Mathematical models*. These models contain a series of equations that represent our theoretical interpretations of real-life situations. For example, epidemiologists may employ mathematical models to predict how a given epidemic might spread in a population, or a physiologist may employ a mathematical model of neuronal membranes to predict the behaviour of action potentials in a neurone. When our theories and related models are sufficiently detailed and well formed, we may use these in simulation research. We now use computers, which are capable of carrying out the complex calculations necessary for the simulation of a real-life situation and predicting numerical outcomes.
3. *'Paper and pencil' models*. Often, our theories are not sufficiently detailed to allow precise numerical predictions. Here the models are 'sketches' of a particular system, defining the key elements of the model and showing how these elements interact to produce various outcomes. Such models enable us to make testable predictions concerning the effects of variables as shown by the following model (Fig. D1.1) of the 'gate control theory' of pain (Melzack & Wall 1965). (Note that for teaching purposes this model is an incomplete and modified representation of the original. If you are interested in understanding pain problems, you should consult Melzack & Wall's original work.)

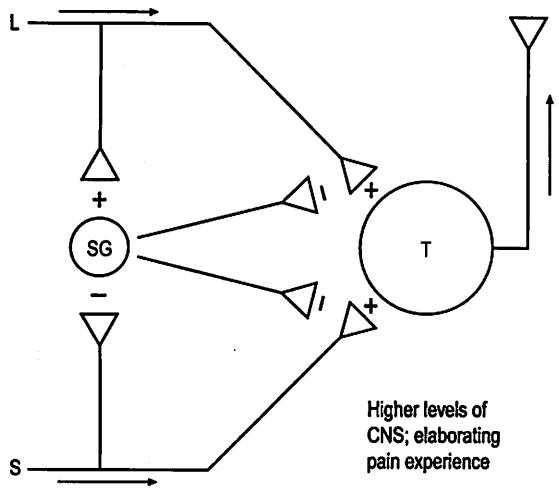


Figure D1.1 • Model of spinal gating of nociceptive input (adapted from Melzack & Wall 1965). See text for details.

Elements of the systems

(a) L: 'Large'-diameter axons of receptors, which convey information concerning mechanical stimuli, such as pressure and vibration.

(b) S: 'Small'-diameter axons of receptors, which convey information concerning noxious (tissue-damaging) stimuli.

(c) SG: Neurones within the substantia gelatinosa (SG) of the spinal cord. These receive converging information from L and S axons. The SG neurones control, or gate through pre-synaptic inhibition, the information flowing through to the T neurones.

(d) T: Transmission neurones in the spinal cord, which receive information from both L and S axons. The pattern of activity of T neurones is projected to higher levels of the central nervous system (CNS), where this information is elaborated into the experience of pain.

(e) Δ: Synapses, which may be either excitatory (+) or inhibitory (−). Excitatory synapses increase, while inhibitory synapses decrease the activity of the post-synaptic neurones.

(f) →: Arrows showing the direction of the information flow; in this case from the periphery towards the spinal cord and, subsequently, to higher levels of the CNS. The activity of large fibres (L) stimulates the gating mechanism (SG) which serves to inhibit nociceptive information conveyed by S axons to the T neurones.

The above model is an attempt to produce a representation of the events which take place in the mammalian nervous system when receiving and processing nociceptive input at a spinal cord level. The 'gate control' model integrates a broad range of research in the neurosciences and has been applied to explain aspects of pain control in clinical settings.

Questions

After this rather prolonged introduction to the model, we are ready to ask questions about how it may be applied to explaining and predicting observations about pain.

1. Use the above model to describe what happens in the nervous system during noxious (tissue-damaging) stimulation.
2. How would we use controlled observations to demonstrate that small (S), rather than large (L), neurones convey nociceptive information? (Assume that we have instrumentation for measuring the activity of single neurones in response to different kinds of stimuli.)
3. Explain how we would use evidence obtained by recording the activity of single neurones to demonstrate that T neurones are in fact involved in nociception.
4. Describe the effects of large (L) and small (S) axons on the activity of SG neurones.
5. Explain the mechanism by which SG neurones function as a 'pain-gating' system.
6. Propose a hypothesis for predicting the effects of selectively damaging T neurones on subsequent pain experience.
7. Propose a hypothesis for predicting the effects of damaging large axons (L) on subsequent pain experience.
8. A virus is identified which damages SG neurones. As a hypothetical case, imagine that people who have this virus report greatly reduced pain sensitivity. What would be the implication of this observation for the validity of the gate control theory of pain?
9. Would you discard the gate control theory on the bases of the hypothetical observations described in question 8? Explain.
10. A clinical technique called transcutaneous electrical nerve stimulation (TENS) for relieving pain involves the gentle peripheral electrical

stimulation of painful areas. Explain in terms of the model how TENS might work to reduce pain. (Hint: TENS does not directly reduce the activity of S axons.)

Answers

1. Small-diameter S axons convey nociceptive information towards spinal cord neurones, including the SG and the T neurones. The SG neurones are inhibited through nociceptive inputs while the T neurones are excited by the stimulation. Information concerning tissue damage is then conveyed by the axons of the T neurones to higher levels of the CNS. There it is elaborated into the experience of pain.
2. By showing that the activity of T neurones is correlated to the levels of peripheral noxious stimulation.
3. Mechanical stimuli – show that S axons are not responsive while L axons change rate of activity. Noxious stimuli – show that S axons change rate of activity while L remain unresponsive.
4. Large axons excite (increase the rate of activity of SG neurones) while small axons inhibit the rate of activity of the SG neurones.
5. The SG neurones inhibit the effect of the S axons on the T neurones. As L axons excite the SG neurones, the action of L fibres is to 'close' the pain gate. The S fibres 'open' the pain gate by inhibiting the SG neurones. In effect, the model is telling us that mechanical stimulation, such as gently scrubbing an injured area, will reduce nociceptive input. Conversely, noxious stimulation seems to maintain the effects of subsequent nociceptive inputs.
6. The T neurones are crucial for transmitting nociceptive information to higher levels of the CNS. Hypothesis: selectively damaging T neurones will reduce or eliminate pain experience following nociceptive stimulation.
7. Damage to L axons will result in reduced excitatory stimulation of the SG neurones, opening the spinal gate to noxious stimulation. Hypothesis: damaging L axons would increase the pain experienced following nociceptive stimulation.
8. As the model proposes that SG neurones are involved in the gating of noxious stimuli, we would predict that damage to the SG neurones would increase pain sensitivity. Reduced pain sensitivity is evidence which would falsify the gate control theory.
9. Probably not; as we discussed in the context of Lakatos' ideas. Theories have a 'protective belt' of assumptions, which means that a single empirical falsification need not result in the rejection of the theory. In this case, we may look for additional effects of the virus; say in destroying CNS tissue involved in elaborating pain experience. However, in the long run, such disconfirmations will lead to discarding the theory.
10. In terms of the model, TENS works by stimulating the L axons, and thereby 'closing' the pain gate, as outlined in questions 4 and 5.

Section **Two**

Research planning

SECTION CONTENTS

The first stage of research involves the detailed planning of the project. The plan for what is to occur in the project is written up in a document called the research protocol. Before the research project may proceed, the protocol is examined by ethics committees and funding bodies to ensure that it conforms with general methodological and ethical principles. The three chapters of Section 2 aim to outline the basic considerations for the successful preparation of a research protocol.

The primary reason for carrying out a research project is to obtain empirical evidence that will advance theory and practice in the health sciences. Before all else, we must be sure that we are asking the right questions, that is, raising issues and problems which are central to progress in contemporary health care. We must convince the critical reader that our aims or the hypotheses which we are attempting to resolve are of central importance. Asking the right research questions depends on being creative, for example identifying previously ignored patterns in the data, or the construction of novel theories that predict new, as yet unobserved phenomena. Of course, we are not suggesting that researchers are geniuses, or that only brilliant, ground-breaking research is funded. Much important research is carried out by perfectly ordinary men and women who are interested in their patients and their problems, who have a good knowledge of their professional practices and who understand research methods.

To justify the research proposal it is necessary to write a 'literature review'. The literature review is a summary and critical evaluation of previous research and theory relevant to the problem we are intending to investigate. In this way the literature review provides both a conceptual background for our proposal and justifies the need for further empirical evidence by identifying 'gaps' in our knowledge.

The proposed research may be descriptive, for example collecting information concerning health needs of a community and/or the impact of illness or injury on a group of patients. In Chapter 3, non-experimental research strategies are described. Non-experimental research which includes both quantitative and qualitative surveys (Section 3) aims to provide a clear picture or description of the

health of individuals, groups or whole communities. Experimental research strategies are appropriate when we are testing hypotheses about the causes of illness, and when we attempt to gain control over extraneous variables which may influence treatment outcomes. The notions of causality and control are central for health research and are explored in Chapter 4. Of course, when we write a research protocol, we must be sure that we select the appropriate research strategy.

No research can proceed unless it is judged to be ethical by an appropriate committee. A research proposal is judged ethical if it conforms with our rules and values concerning caregiving. These rules and values are made explicit in documents representing the standards of professional groups and of institutions (e.g. hospitals, universities and research councils). The research protocol must be described in sufficient detail so that a decision can be made as to whether any harm might occur to participants in the research project.

Economic considerations, or the availability of resources, also have a strong influence on research planning. For example, you may have designed a qualitative research project which involves 100 'in-depth' interviews with persons suffering from a disorder. Say you have only one year and a very limited amount of money to complete your project. You would be advised either to reduce your sample size or, if this is not possible, change your topic. Ethics or funding committees will only approve projects which can feasibly be completed with the available resources.

The way in which samples are selected is discussed in Chapter 3. Selection of an appropriate sample is crucial for the generalizability (or external validity) of your findings. Our aim is to select a representative sample of the population but this may not always be possible in health sciences research. We need a sample size which is sufficiently large to identify the phenomena in which we are interested, but not too large or we are simply wasting resources.

Thus, research planning is a process by which we transform our ideas into well-planned, ethical and economically feasible projects, as described in a research protocol.

Chapter Two

2

Research planning

CHAPTER CONTENTS

Introduction

Before the actual collection of information or data begins, researchers generally invest considerable time and effort in the planning of their investigations. The first step is to select and to justify the research problem. The second step is to transform the problem into clear researchable aims and research questions and, in the case of quantitative research, hypotheses are also usually specified. Research planning includes the selection of an appropriate research strategy for providing the required evidence to answer the research question. The selection of the appropriate research strategy may depend not only on the research questions being asked, but also on ethical issues and resource constraints that may define the scope and form of the investigation.

The specific aims of this chapter are to:

1. Discuss how research questions are selected and justified and derived from research aims.
2. Broadly outline research strategies available for specific types of investigations.
3. Discuss the ethical and economic constraints on the planning and implementation of research.

Sources of research questions

Advances in health research depend on the identification of questions and problems that promote

the development of more powerful theories of health, illness and disability, and then devising more effective ways of assessing, treating and preventing health problems. This advancement depends very much on researchers asking the 'right' questions and identifying solvable problems and then applying sound investigation methods to these questions and problems.

The formulation of research questions depends on the expertise of researchers who have a combination of theoretical knowledge and practical experience for identifying problems and asking the right questions. The professional background of researchers is an important consideration, as one's educational background and practices will influence what are seen as theoretically and professionally interesting problems.

The formulation of even apparently simple research questions is often the culmination of intensive preliminary observations, spending long hours in the library reading through and critically analysing related research, discussing the issues and thinking through a range of issues.

The health care setting in which the researcher works may also have a strong influence on the formulation of research questions. For example, it is an important consideration if the researcher works in a laboratory, a hospital, a community clinic or in private practice. These settings will often influence how the patients' problems are perceived, and how researchers define professionally relevant solutions to these problems.

In many areas of health care, including areas such as rehabilitation, community health and neurosciences, researchers sharing similar interests may join together in multidisciplinary research teams. Because of the broad range of expertise and perspectives, such teams may be in a position to formulate interesting research questions that can be simultaneously addressed from these different perspectives.

The formulation of research questions

The formulation of research questions is a creative act. It draws upon consideration of previous relevant work in the area under study. This work may be reviewed and the review may give rise to questions that have not been answered by previous work. A literature review is the usual way in which knowledge gaps are identified. The purpose of the review is to answer the questions 'What do we know?' and 'What do we not know?'

Inspiration for research questions often stems from personal experience. Exposure to an issue or problems may raise questions suitable for a research study. Most health practitioners constantly pose questions like 'What is the best and most effective method of treating this condition?' and 'How can we ensure that patients or clients get what they want from this service?'

The justification of research questions

When researchers call on public funds to conduct their studies, they must explain in what way their intended study will contribute to scientific knowledge and/or clinical practice. In 'pure' or 'basic' scientific research, a proposed investigation is justified in terms of its potential contribution to existing knowledge. In the context of 'applied' health research, the investigator may be required to demonstrate that the research will in some way contribute to the improved practice of health care, and benefit patients.

Before embarking on the design and conduct of a research project, the investigator must review previous work and publications relevant to the aims of the intended project. This process is essential, both for providing the appropriate background and context for the investigation, and for justifying the investigation in contributing to existing knowledge. It is a waste of money to duplicate very similar research (although deliberate replication of previous studies may be a legitimate activity if there are uncertainties about the validity of a previous study).

Literature searching may be carried out at appropriate research centres or libraries where scientific and professional journals are stored and, of course, by using the Internet. There are powerful web-based search methods that can simplify the search.

The most widely used search tool for health research publications is probably PubMed, the web-based Public Medline (see http://www.ncbi.nlm.nih.gov/entrez/). Professional library staff can help you to locate the relevant literature. However, the critical evaluation of the literature depends on the application of research methods for the identification of controversies or 'gaps' in the available evidence.

Formulation of research aims

Hypotheses in quantitative research

Quantitative research is normally structured so as to test a research hypothesis. Hypotheses are propositions about relationships between variables or differences between groups that are to be tested. Hypotheses may be concerned with relationships between observations or variables (for example, 'Is there an association between level of exercise and annual health care expenditure?') or differences between groups (for example, 'Do patients treated under therapy x exhibit greater improvement than those treated under therapy y?'). A *variable* is a property that varies, for example the room temperature, the ages of the patients, or their improvement on a clinical measurement scale.

Some quantitative research projects do not have a hypothesis to be tested in any formal sense. For example, if you are assessing the health needs of a local community, there need not be any specific expectation or hypothesis to be tested. The purpose of the research may be simply to describe accurately the characteristics of the study sample and target population. The goal is not to test a specific hypothesis. This does not mean the research is deficient; it just means it has a different objective from other types of quantitative research projects. So hypothesis testing is a characteristic of many, but not all, quantitative projects.

Qualitative research does not usually test hypotheses. It is concerned with understanding the personal meanings and interpretations of people concerning specified issues. However, most qualitative researchers work with specified research aims and questions to guide their investigation.

The formulation of adequate research aims generally involves a process during which the researcher gradually refines a broadly based issue into more specific operationally defined statements. This process must take into account the logical requirements of designs for information and data collection and ethical standards for conducting health research. In addition, the research aims must inform a realistic data collection procedure which can be supported by the resources available for researchers. We will examine these issues in the rest of this chapter.

Research strategies

Research planning also involves the selection or formulation of research strategies. Research strategies are established procedures for designing and executing research. The research strategies that are outlined in this book include experimental and quasi-experimental strategies, single case research, surveys and qualitative research. Before detailed discussion of the differences between these strategies is undertaken in the following sections, it is instructive to compare the basic structures of survey research and experimental research. As we shall see in subsequent sections, these are fundamentally different ways of conducting research.

Survey research strategies

The three essential steps involved in survey research strategies are shown in Figure 2.1. You will note in Figure 2.1 that the researcher defines the population of interest, selects the cases to be studied and then observes or studies them, generating the data. There is no intended active intervention in the situation by the researcher. Indeed, intervention that would change the phenomenon being studied is discouraged and avoided. In this type of research strategy there is a constant tension between the requirements for close and detailed observation (proximity to the phenomenon under study) and the possibility that the behaviour of the research participants under study may change because they are being studied. There are techniques for dealing with this problem, such as the use of unobtrusive

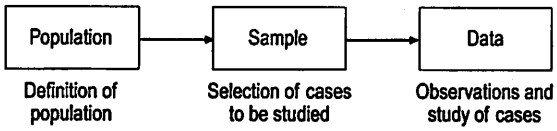


Figure 2.1 • Structure of a survey research strategy.

measures and participant observation, which will be discussed in subsequent chapters.

Experimental research strategies

Experimental research strategies involve a sequence of steps, as illustrated in Figure 2.2. As with the survey research strategy, cases are selected for study from a defined population. However, these cases are then assigned (generally by chance or random method) to a treatment (intervention) or non-treatment (non-intervention) group, and then given the intervention or, in clinical contexts, the treatment. The term 'treatment' is used here in the sense that the subject or event is being systematically influenced or manipulated by the investigator; that is, in a broader sense than with medical or physical treatment. For this reason the broader term 'intervention' is often used in place of the term 'treatment'. The intervention and non-intervention groups are then observed and compared. If everything went according to plan, any differences in the results for the intervention and non-intervention groups should be a result of the effects of the intervention. This type of design is often called a randomized controlled (sometimes clinical) trial or RCT. In some review systems such as the Cochrane Collaboration system, the RCT is considered to be the 'best' level of evidence to establish the effectiveness of an intervention.

Thus, there is quite a difference in the structure of experimental and survey (non-experimental) and qualitative research strategies. Cook & Campbell (1979) proposed a separate class of '*quasi-experimental*' strategies (which are really just tightly structured non-experimental strategies) that we will also cover later in this book.

The decision of the investigator to adopt specific research strategies depends on the phenomenon being studied and the specific question type of research (causal or descriptive) being investigated.

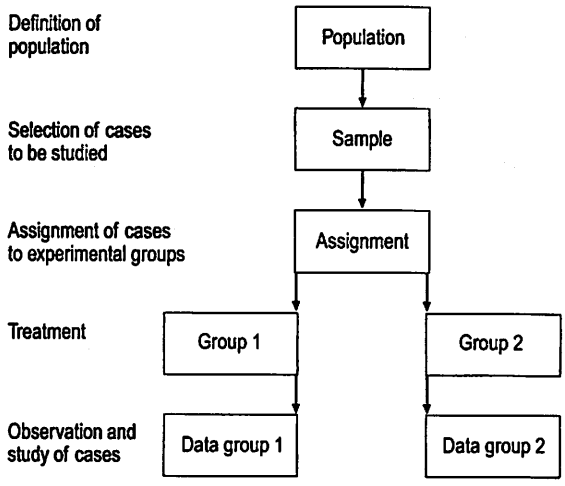


Figure 2.2 • Experimental research strategies.

Research planning: ethical considerations

In health research, where the health and lives of people participating in the study may be at stake, ethical considerations play a key role in research planning and implementation. For instance, consider the question of whether cigarette smoking has negative impacts on health. One way of investigating this question would be through experiments, which are generally considered to be the design best suited to establishing causal relationships. However, it would be ethically unacceptable to have a number of people randomly assigned to a group involved in long-term, heavy cigarette smoking as this would very likely damage their health. Clearly, to meet ethical criteria, evidence must come from studies in which investigators observe the effects of smoking in persons who have themselves chosen to smoke.

A research process is judged to be *ethical* by the extent to which it conforms to or complies with the set of standards or conventions in the context in which the research is to be carried out. In different countries, the various medical research councils

have developed detailed statements concerning these standards. In the United Kingdom the Medical Research Council has a series of statements and publications available at its web-site (http://www.mrc.ac.uk), as does the Australian National Health and Medical Research Council on its web-site at http://www.nhmrc.gov.au. Canada has the statements by the Interagency Advisory Panel on Research Ethics (http://www.pre.ethics.gc.ca). In the United States there is a proliferation of bodies with research ethics guidelines, mostly based upon the 1979 work by the Belmont group. All of these guidelines draw upon the pioneering 1964 Helsinki declaration from the World Medical Association.

Most health services and higher education institutions now have their own ethics committees to oversee health research involving humans and/or animals. Ethics committees, along with the researcher, are responsible for interpreting and enforcing ethical standards. It should be noted, however, that even when a given research proposal is judged to be ethical, it may not be seen as *moral* by specific individuals or groups in a community. A good example is animal experimentation when the research involves painful or stressful procedures such as surgery. According to the value systems of some persons and groups, such work may be considered to be intrinsically immoral. From this viewpoint, there may be no acceptable justification for inflicting suffering on animals, even when the results may be beneficial for humans. Others may argue that such procedures are morally justified, given that they are necessary for the advancement of the biological and medical sciences and provided that reasonable steps are taken to limit the suffering of the animals. There are no absolute solutions for such controversies as they involve human values, but discussing these issues in public may help to establish a degree of consensus in the community and guidance for the ethical decisions of researchers.

It is beyond the scope of this discussion to examine in full the ethics of health research, but the following issues are central to making decisions in these areas:

1. *Benefits.* Who is to benefit from the research? One likes to think that 'humanity', 'the participants', or 'health science' are going to be the beneficiaries of health research. A slightly more cynical analysis points to the investigators as having the most to gain, at least in the short term. After all, given a successful outcome for the research project, they stand to satisfy their curiosity, improve their career prospects or to raise their esteem before colleagues. In practice, benefits accruing to the investigator, the participants and health science have to be carefully weighed in relation to the conduct of any project. The protection of the rights and welfare of the participants is the primary consideration of research ethics.
2. *Informed consent.* Informed consent by the participants is a necessary condition for ethically acceptable research. This means that all the risks involved in the investigation must be explained to prospective participants, as well as the possible benefits, and they must decide about their participation or non-participation in an informed manner. Honest explanation of the procedures to the participant takes considerable skill, as disclosing the design of the study may influence the expectations and behaviour of the participants. Also, special care needs to be taken when the participants are in some way limited in their ability to understand the risks; for example, people who are intellectually disabled, or confused under the influence of drugs. Informed consent implies a freedom of choice: participants must feel confident that refusal to participate will not prejudice their subsequent clinical treatment.
3. *Protection of participants.* It is the investigator's responsibility to minimise the chances of long-term deleterious effects to subjects. Dangers can arise from administering new interventions or treatments with unknown side effects, denying interventions of known effectiveness or using invasive assessment techniques. The welfare of laboratory animals must not be ignored. There are nowadays legally enforceable constraints in most jurisdictions on the conditions for using animals in research.
4. *Minimizing discomfort.* The investigator must also minimize even short-term pain, anxiety, discomfort or embarrassment involved in an investigation, especially if it is not part of routine therapeutic practice. This is an important issue, as some therapists and researchers take a mistakenly 'proprietary' view of their patients, imagining that they should put up with some discomfort for the good of medical science. In the past, patients in public wards of hospitals and long-term prisoners have been most vulnerable to questionable ethical practices.
5. *Privacy.* Health care research may involve collecting information that can lead to the embarrassment

or stigmatization of the participants. The identity of participants must be protected; for instance, by 'coding' their real names so that they may not be identified even inadvertently. In qualitative research where the full text of comments may be reported, it is important that this text does not inadvertently identify people through the reporting of identifiable scenarios or names.

6. *Community values.* Some research, such as that involving the assessment of sexual functioning or the investigation of human fetuses, calls for very sensitive planning. Anatomical research involving the use of human cadaver material must be planned to ensure respectful handling and disposal. Even if the investigator does not agree, the values and taboos of groups in the community must be taken into consideration when planning the investigation.
7. *Conservation of resources.* A fundamental value is that the time and effort of researchers and subjects and the resources of the community should not be wasted on a badly planned investigation. Of course, there can never be a guarantee that an investigation will produce clinically useful results. However, some poorly designed projects are doomed to failure even before the data collection begins, and lead to confusion and controversy in the professional literature. Therefore, the appropriate use of high-quality and sound research methods is not only useful for solving problems, but constitutes an ethical necessity. Many ethics committees also vet projects for scientific merit before the project is approved.

A responsible investigator is required to take into consideration ethical principles and to plan research projects accordingly so that no, or the minimum possible, harm is caused. Therefore, the health researcher must use considerable ingenuity in designing valid investigations, while maintaining ethical values. One of the roles of ethics committees is to guide investigators on complex issues, and to ensure that research is conducted in accordance with accepted community principles.

Selection of research strategies: economic considerations

Selection of appropriate research strategies can also be influenced by the resources available to investigators. Some projects involving the evaluation of the safety and effectiveness of new drugs, or the identification of risk factors in cardiac disease, have taken tens of millions of dollars to finance. Research planning takes into account economic issues, such as:

1. *Availability of participants.* Here the investigator has to consider if enough people can participate in the project under the required conditions. Attention has to be paid to issues such as the frequency of the disorder in the community, problems in identifying participants and their level of voluntary participation. As will be shown in Chapter 3, selection of participants is crucial for the validity of a study.
2. *Availability of equipment.* Some equipment is costly to acquire or to operate (for example, CT (computed tomography) scanners, biochemical assays, experimental drugs). Research planning should take these expenses into account.
3. *Availability of expertise.* Specialized assessment techniques and the administration of clinical treatments require professional expertise. If these are beyond the investigator's competence or if the experimental design requires an unbiased, 'blind' therapist or assessor (see Ch. 5), there must be means for securing them.
4. *Availability of time.* Research projects have a tendency to take considerably longer than an inexperienced researcher might expect, due to the erratic and at times disastrous workings of Murphy's law: 'If anything can go wrong, it will'. Equipment breaks down and needs to be repaired, subjects 'disappear', collaborators might not deliver services promised. Research planning should take into account possible problems and how these might affect the time scale of investigations. This is a particularly important issue for postgraduate students.

The availability of resources strongly influences the scope of the research programme and also the research strategy selected by the investigator. A research project must be shown to be feasible from a cost viewpoint before it is initiated. It is only after the ethical constraints and economic resources have been evaluated that the researchers' questions can be transformed into clearly defined hypotheses or aims.

It is important when conducting research that the identities of all supporting agencies and individuals are disclosed to the participants in

order that their agreement to participate is fully informed.

Steps in research planning

The following steps should be considered in quantitative research before actual data collection begins. The term 'data' refers to the set of observations recorded during an investigation.

1. Identification of the research problem.
2. Retrieval and critical evaluation of relevant literature for justifying and giving the context of the research problem.
3. Formulation of precise research aims and questions in the light of:
   (a) the information to be collected
   (b) the appropriate research strategies
   (c) the ethical considerations required to protect participants and the community
   (d) the projected cost of the project.
4. In addition, as to be discussed in this book, research planning takes into account:
   (a) how the participants are to be selected
   (b) the design of the investigation
   (c) how the data are to be collected
   (d) how the data are to be analysed.

The above points are written up in the form of a *research protocol*. This protocol is submitted to supervisors, colleagues or ethical committees, for scrutiny of methodological and ethical problems. It is only when the protocol is acceptable to all these parties that the actual data collection begins. It is often desirable to carry out a small-scale preliminary study, called a 'pilot'. This is an economical way of identifying and eliminating potential problems in the large-scale investigation.

Summary

The planning of a research project requires the transformation of a vague question or problem into clearly stated aims or questions. To achieve this, the researcher should review the relevant literature and evaluate ethical considerations and economic constraints in conducting the investigation. Next, an appropriate research strategy needs to be selected. On the basis of the above, the researcher is in a position to state precisely the aims or hypotheses being investigated.

Before data collection begins, a protocol for the complete investigation is scrutinized by colleagues to correct methodological problems, or to prevent unethical research.

Self-assessment

Explain the meaning of the following terms:

ethics
informed consent
literature review
Murphy's law
non-experimental study
pilot study
research strategy
variable

True or false

1. Scientific research always involves the testing of hypotheses.
2. A sample is a subset of the data.
3. The population being studied is defined after the sample has been selected.
4. A research protocol is a summary of the data obtained in an investigation.
5. A pilot is a small-scale research programme to demonstrate the feasibility of an investigation.
6. Human cadavers are not sentient beings, therefore their treatment falls outside the scope of medical ethics.
7. Planning quantitative research includes specifying the variables to be studied and how these are to be measured.
8. Abstract constructs, like 'intelligence' or 'motivation', cannot have empirical referents.
9. A literature review is necessary for providing the appropriate background and rationale for an investigation.
10. Well-established data collection techniques are called research findings.
11. A non-experimental design involves the assignment of subjects to treatment groups.
12. Both experimental and non-experimental designs involve the drawing of an appropriate sample from the population being studied.

13. An anthropologist studying how ethnic groups relate to illness would tend to adopt a quantitative research strategy.
14. The clear explanation of the risks involved in a research project is a necessary condition for obtaining informed consent from a subject.
15. The use of correct research methods does not constitute an ethical necessity.
16. A hypothesis is a proposition about the relationship existing between variables or prediction of differences between groups.
17. Some scientific research projects do not involve the testing of hypotheses.
18. Existing theories are the only appropriate sources for experimental hypotheses.
19. In a non-experimental study there is no need to pay attention to sampling.
20. Experimental designs involve the assignment of the sampled subjects into treatment groups.
21. Experimental designs are more appropriate than non-experimental designs for demonstrating causal relationships.
22. The subjects participating in medical research are generally the most likely beneficiaries of the project.
23. Informed consent of human subjects is necessary even when the investigation is not apparently embarrassing or dangerous.
24. It is unethical to study human sexual behaviour because people find it embarrassing to serve as subjects.
25. Medical researchers should expect that patients who receive free hospital treatment should participate in medical experiments.
26. The correct use of research methods in conducting an investigation is useful, but not an ethical necessity.
27. A well-designed, relevant research project will always be financed, regardless of the expense.
28. A variable is a property or attribute which varies from subject to subject.

Multiple choice

1. The aim of research planning is to:
   a generate appropriate aims or clear-cut research hypotheses
   b select an appropriate research strategy
   c identify possible ethical or economic limitations in conducting the investigation
   d *a* and *b*
   e all of the above.
2. A 'literature review':
   a is a list of research publications relevant to an investigation
   b should discredit research findings which are inconsistent with the hypothesis
   c should include only findings which directly support the hypothesis being investigated
   d should be a critical review of findings relevant to an investigation.
3. Which of the following is common to both experimental and non-experimental research strategies?
   a random assignment
   b selection of cases to be studied
   c experimental hypothesis
   d participant observation
   e field research.
4. Which of the following is unique to the experimental research strategy?
   a random assignment
   b selection of cases to be studied
   c definition of population
   d participant observation
   e field research.
5. The availability of resources to conduct an investigation:
   a is not the concern of a scientific researcher
   b has an influence on the scope and design of the investigation
   c is determined by Murphy's law
   d none of the above.
6. 32 degrees Centigrade is an example of:
   a a variable
   b the value of a variable
   c a ratio
   d a null hypothesis.
7. A variable is:
   a a property which can take different values across different individuals
   b a property which can take different values within an individual
   c both *a* and *b*
   d neither *a* nor *b*.
8. Which of the following ethical requirements is the *most* difficult to satisfy when conducting a placebo-controlled randomized clinical trial?
   a Ensuring the privacy of the research participants.
   b Obtaining informed consent from all research participants.

c Protecting research participants from unnecessary risks.

d Providing the best available treatment for all research participants.

9. Which of the following represents the most explicitly formulated research aim?
   a The aim of the present project was to investigate if health care workers in Gotham City were satisfied with their pay and career prospects.
   b The aim of the present project was to investigate if emotionally disturbed persons living in Gotham City received adequate medical care.
   c The aim of the present project was to identify the reasons why health professionals in Gotham City leave their professions, and turn to other types of employment.
   d The aim of the present project was to identify the proportion of teenagers in Gotham City smoking more than 10 cigarettes a day.
10. A research protocol should make explicit:
   a the justification for undertaking the research project
   b the way in which the data are to be collected
   c the empirical evidence provided by the investigation
   d *a* and *b*.
11. Which of the following represents a non-researchable problem (in the context of the scientific method)?
   a Is aspirin preferable to steroids in reducing the symptoms of arthritis?
   b Does the use of alcohol result in greater levels of brain damage than the use of marijuana?
   c Are painful experiments involving animals justified if they lead to benefits to patients?
   d Do people coming from a low socioeconomic class receive poorer health care than people from a higher class?

Chapter Three

3

Sampling methods and external validity

CHAPTER CONTENTS

Introduction

Research in the health sciences usually involves the collection of information from a sample of participants, rather than on the entire population in which the investigator is interested. Studies that involve an entire population or group are called *census* studies but these are relatively rare and generally expensive to perform. A sample drawn from the target population is studied because it is usually impossible or too costly to study entire populations. For instance, when individuals who have conditions such as diabetes, cerebral palsy or emphysema are being studied, it is not possible to study everyone because of the large size of such populations. Also because many people do not seek treatment or may be wrongly diagnosed, we may not be able to identify all members of the entire population in order to study them. Therefore, in most research, the researcher studies a subset or sample of the target population, and then attempts to generalize the findings to the population from which the participants were drawn. This general principle applies to both qualitative and quantitative studies.

The aim of this chapter is to examine ways that samples can be drawn to permit the investigator to make valid generalizations from the study sample to the target population. We will also consider the question of generalizing the findings

of an investigation to other samples and situations. This is referred to as 'external validity' or 'generalizability'.

The specific aims of this chapter are to:

1. Define what is meant by sampling and representative samples.
2. Outline the relative advantages and disadvantages of commonly used sampling methods.
3. Discuss the relationship between sampling error and sample size.
4. Examine the concept of external validity for generalizing research findings to other settings.

What is sampled in a study

While this chapter will focus on the selection of the research participants in a study, many other things are also selected or sampled. These include:

1. the information to be collected by the researcher
2. the procedures used for the collection of the information
3. where the research is conducted, e.g. in a field setting or in a structured research setting
4. the clinicians and researchers who are involved.

Many researchers focus on the selection of the research participants as the key or only issue in maximizing the generalizability of their research, and they do not pay enough attention to the other factors they are sampling or the context in which the research is being conducted. It is not at all unusual to see studies that employ large and sophisticated participant samples, yet with only one or two highly selected clinicians involved in the research in perhaps only one health setting. While the study sample may be highly representative, the context in which the research is conducted may not be and it may be that the researchers and clinicians involved in the study have a particularly unusual or idiosyncratic approach to their work that is not reflective of others. It is our contention that in many qualitative studies there is strong consideration of the research context and its impact upon the research findings. However, there is often less emphasis upon sampling of research participants. This impacts upon the ability to generalize the findings more broadly from the actual research participants to other groups.

Basic issues in sampling

As we have discussed, often, because of the numbers involved, it is not within the resources of the researcher to study the whole target population. In any event, in most situations it would be wasteful to study all of the population. If a sample is representative, one can generalize validly from the sample's results to the population without going to the expense of studying everyone.

The population is the target group of individuals or cases in which the researcher is interested. Examples of valid study populations include: all English women under 25; all children with diagnosed spina bifida in the state of Alberta; all the students at a particular Australian college. The researcher defines the population to which he/she wishes to generalize. Note that a population need not consist of human or animal subjects. Objects or events can also be sampled, as shown in Table 3.1.

As shown in Table 3.1, a population is an entire set of persons, objects or events which the researcher intends to study. A sample is a subset of the population. Sampling involves the selection of the sample from the population.

Representative samples

There is a variety of different ways by which one can select the sample from the population. These are called sampling methods.

The ultimate aim of all sampling methods is to draw a representative sample from the population. The advantage of a representative sample is clear: one can confidently generalize from a representative sample to the rest of the population without having to take the trouble of studying the rest of the population. If the sample is biased (not representative of the population) one can generalize less validly from the sample to the population. This might lead to quite incorrect conclusions or inferences about the population. This would mean that the results obtained in the study would not necessarily generalize to other studies using the same population. Figure 3.1 illustrates the concept of a representative sample.

Table 3.1 Examples of populations and samples

Population	Possible sample selection
All working podiatrists in a state	50 podiatrists selected for a study of job satisfaction
All working pathologists in a state	25 pathologists selected for detailed tax evaluation by an inspector
The temperature of a patient during a 24-hour period	Hourly measurements of the patient's temperature recorded by staff
Stuttering in a child's speech	Number of stutters made during 5 minutes of reading a standard piece of material
All patients in a state with frontal lobe damage	30 patients with frontal lobe damage selected for evaluating a rehabilitation programme
All surgical gauzes held by a given hospital	10 gauzes selected by a bacteriologist to test for sterility

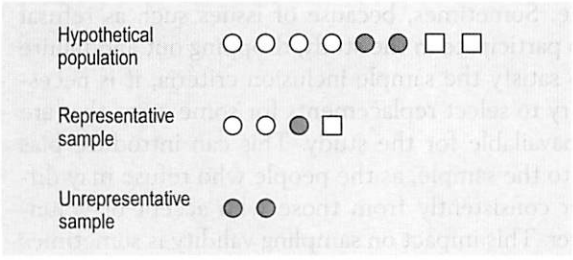


Figure 3.1 • A simple representative sample.

Figure 3.1 illustrates a hypothetical population composed of three different types of study participants or categories of participants. A representative sample is a precise miniaturized representation of the population. An unrepresentative or biased sample does not adequately represent the key groups or characteristics in the population, and this may lead to mistaken conclusions about the state of the population.

The selection of the appropriate sampling method depends upon the aims and resources of the researchers. For instance, if someone is designing a very expensive health or social welfare programme on the basis of a survey of clients' needs, it is imperative that the researcher uses a good sampling method and obtains a representative sample of the clients, so that appropriate conclusions may be reached about the population. Good sampling methods are somewhat more expensive and more difficult to implement than poor methods but they are worth it. The main sampling methods used in health research are incidental and random sampling.

Incidental samples

Incidental sampling is the cheapest, easiest and most commonly used sampling method in clinical studies. It involves the selection of the most accessible and available members of the target population. For example, a researcher who stands in the middle of a city street and quizzes people about their health status is practising incidental sampling. However, it is quite likely that this sample would not be representative of the general voting population. There would probably be an over-representation of businessmen and white-collar workers, and an under-representation of factory workers and housewives. The sample is likely to be unrepresentative and biased.

A further example of incidental sampling might involve a researcher surveying the needs of a group of spina bifida children at a local community health centre. Their measured needs may be representative of those of other spina bifida children, but then again they may not if these children are not typical of the wider population of children with spina bifida.

Thus, incidental sampling is cheap and easy to implement but may give a biased sample that is not representative of the population.

Quota sampling

Sometimes it is known in advance that there are important subgroups within the population that

Table 3.2 Distribution of percentages of gender and occupational variables in the general population

	Blue collar	White collar	Not employed	Total
Male	19	21	9	49
Female	15	15	21	51
Total	34	36	30	100

need to be included in the sample. Two such important groups within the human population are males and females. Further, it is known that they occur in the ratio of approximately 49:51 in the general population. Our researcher might decide that it is very important that the sexes are proportionally represented in the sample. Thus, the researcher would set two quotas (of 49 male and 51 female respondents in a sample of 100) and sample accordingly in a city street. This is still a form of incidental sampling but has some significant advantages over simple incidental sampling because the study sample's composition on this key demographic variable is guaranteed to match that of the target population.

More sophisticated examples involving more than two groups can be accommodated as shown in Table 3.2. We can see from Table 3.2 that if our sample were to be representative regarding both sex and occupational status, in a sample of 100 people we would need 19 blue-collar males, 15 blue-collar females, and so on.

Quota sampling still has a number of shortcomings: before it can be used, one has to know which population groups are likely to be important to a particular question and the exact proportions of the various groups in the population. Sometimes we may not know these proportions. Also, the members of the sample within the quotas are still incidentally chosen. The blue-collar males, for example, selected in a city centre on a weekday may still be quite different from those working elsewhere. However, quota sampling is better than simple incidental sampling.

Random and systematic samples

Random sampling

This is one of the best but probably more expensive sampling methods to implement in drawing a study sample. A random sample is one in which all members of the population have an equal chance of selection. Thus a random sample is more likely to be representative of the relevant population than an incidental sample.

The procedure for drawing a random sample involves:

1. the construction of a list of all members of the population
2. using a method such as dice, coins, hat or random number tables to select randomly from the list the number of members required for the sample.

A simple example of a random sample is provided by a common raffle, where names on (preferably!) equal-size papers are put in a hat, shaken and selected 'blind'. Many national and state lotteries use numbered balls that are drawn randomly from a barrel. Another way to draw a random sample is to construct a list of all the members of the population and assign a number to each element. Then a table of random numbers, generated by a computer, could be used to select a random sample.

Cases that are selected from a list using a random selection method constitute a random sample. Sometimes, because of issues such as refusal to participate in the study, dropping out and failure to satisfy the sample inclusion criteria, it is necessary to select replacements for some cases that are unavailable for the study. This can introduce bias into the sample, as the people who refuse may differ consistently from those who accept or volunteer. This impact on sampling validity is sometimes called the volunteer effect.

However, random sampling methods have a number of important advantages over incidental or non-random methods:

- Because the exact sizes of the sample and population are known, it is possible to estimate exactly how representative the sample is; that is, the size of the sampling error. This cannot usually be done with non-random sampling methods.
- Because random samples are usually more representative than non-random samples, the

sample size needed for good representation of the population is smaller.

The major disadvantages of random sampling methods are:

- The researcher needs to be able to list every member of the population. Often this is impossible because the full extent of the population is not known. For example, it would be very difficult to sample randomly from the population of Canadians with coronary disease, because no such list exists. Even if it did exist it would be constantly changing.
- Cost. It is usually easier to use conveniently available groups. Random sampling usually involves considerable planning and expense, especially with large populations.

Stratified random sampling

This involves the same approach as quota sampling with set quotas from specific subgroups, except that each quota is filled by randomly sampling from each subgroup, rather than sampling incidentally. For example, if one was drawing a sample stratified with respect to sex, one would prepare a list of all females and all males in the target population and then sample randomly from these lists with the numbers of each group in the sample corresponding to the population proportions. These groups collectively are called the strata.

The advantages of stratified random sampling are:

- All the important groups are proportionally represented.
- The exact representativeness of the sample is known. This has important statistical ramifications.

The disadvantages are:

- A list of all members of the population, their characteristics and the proportions of the important groups within the population need to be known.
- Cost.
- The gain in sample accuracy is usually small in comparison to simple random sampling.

Area sampling

In area sampling, one samples on the basis of location of cases. For example, on the basis of census data the investigator may select several areas in a city or county with known characteristics, such as high or low unemployment rates. The areas could then be further divided into specific streets and the occupants of, say, every third house contacted for participation in the study. In other words, the locations are randomly selected and then one interviews the occupants of those locations. This can be a very effective, cheap method of sampling in social surveys. It does not require a list of the individual members of the population, merely the location where they live. Recent research has shown that where people live can be an important factor in their health status and their use of health services. The study of the impact of where people live upon health and social issues is sometimes called (geo-) spatial analysis.

Systematic sampling

This involves working through a list of the population and choosing, say, every 10th or 20th case for inclusion in the sample. It is not a truly random technique but it will usually give a representative sample. It is based on the (usually justified) supposition that cases are not added to the list in a systematic way which coincides with the sampling system. Provided a list of cases is available, systematic sampling is an easy and convenient sampling method. In clinical practice we are using systematic sampling when, for instance, we measure temperature and blood pressure every hour.

Sample size

One of the most poorly understood aspects of sampling is the number of cases that should be included in a study sample. The whole issue of sampling is conceptualized somewhat differently for qualitative and quantitative studies, although the issue of to what extent the results of a study are generalizable is the same whatever the study design.

It is obvious that in one sense the more cases selected the merrier (or the better the sample representativeness), but the costs associated with the sampling and data collection must be weighed up against the greater generalizability that is generally associated with larger samples. Also, some studies may involve discomfort, pain or even danger

to patients or laboratory animals. Therefore, in this case it is ethically and logistically desirable to ensure that no more than the bare minimum of subjects is used to achieve the desired sample accuracy. Health researchers 'walk a tightrope' in deciding the optimum sample sizes for their studies. However, there are some principles available to guide the researcher.

First of all, let us say quite definitely that there is no magic number that we can point to as an 'optimum' sample size. Rather, the optimum sample size depends on the characteristics of an investigation in the context of which the sample is drawn. In general, the optimum sample size is one which is adequate for making correct generalizations from the sample to the target population. Let us discuss these issues by introducing the concept of sampling error.

Sampling error

In the quantitative research framework, sampling error is the discrepancy between the true population parameter and the sample statistic. For example, if I happen to know from census data that the actual average age of males in a district is 35 years and the average age of a sample of males I have surveyed from the district is 30 years, then I have a sampling error of 5 years. However, if we do not know the actual population parameters (which is commonly the case), we can only estimate the probable sampling error.

Sampling error is related to sample size by the following relationship:

$$\text{Sampling error} \propto \frac{1}{\sqrt{n}}$$

What the above equation claims is that the greater the sample size (n), the smaller the probability sampling error. In fact, the sampling error in a study sample is inversely proportional to the square root of the sample size.

From this relationship it can be seen that doubling the sample size would only result in a reduction of the error by a factor of the square root of 2 (1.414). Similarly, a ninefold increase in sample size would result in only a threefold reduction in the sampling error. Figure 3.2 illustrates this point by showing the graphical relationship between probable sampling error and sample size.

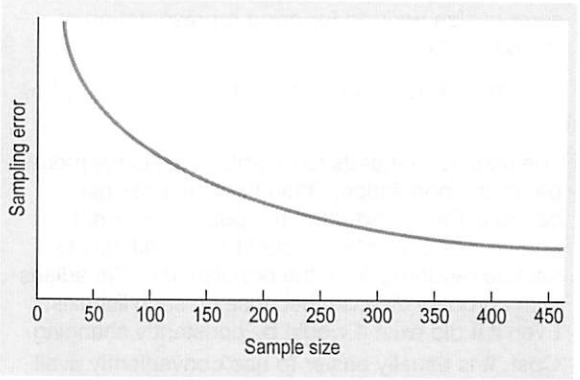


Figure 3.2 • The relationship between sample size and sampling error. The scaling but not the form of the curve will alter with the variability of the data.

It can be seen from this graph that not much is gained from a sample size of over, say, n_1. Yet the cost of the sampling and data collection can be very high with large numbers (such as n_2), for relatively little gain in reducing sampling error. In some research situations, even large probable sampling errors have relatively little potential influence on our decisions. In such situations, we can live with a relatively small sample size. In other situations, we need large samples to justify our confidence in the truth of our decisions.

As an illustration, suppose that we are attempting to predict the outcome of an election fought between two political parties, A and B. A representative sample of 100 respondents is polled before the election. Say the outcome is as follows:

Intends to vote for A	Intends to vote for B	Estimated sampling error
25%	75%	10%

In this instance, the estimated sampling error is very small in relation to the size of the effect (that is, the difference between the percentage of intended votes for the two parties). We can predict confidently, assuming that the respondents

were truthful and don't change their minds before the elections, that political party B will romp into government. Increasing sample size, say to 10000, would enormously increase the cost of the survey. The corresponding reduction in sampling error would not justify this cost, as we would still come to the same conclusion. However, say the following sample statistics were obtained in the pre-election poll of $n = 100$ respondents:

Intends to vote for A	Intends to vote for B	Estimated sampling error
48%	58%	10%

Now the same level of estimated sampling error is too large, in relation to the apparent size of the effect, to make a decision concerning which party is likely to win the election. The poll would have to be repeated using a substantially increased sample size to reduce the sampling error.

The example illustrates the notion that the adequacy of the sample size is affected by the specific situation that we are trying to research. A sample size which is adequate in one situation may be inadequate in another. One benefit of a previous pilot study is that it allows the researcher to estimate the size of the phenomenon under study, and thereby make a more educated guess concerning the sample size required. The concept of statistical 'power' or the ability to detect real effects or differences is discussed later in this text.

Sampling issues in qualitative research

Qualitative research shares with quantitative research a concern with the extent to which findings from one study can be generalized to other settings and people. Therefore the representativeness of the information obtained from the sample is important in qualitative research. However, there are differences in the way in which a representative sample is conceptualized in qualitative research.

The key difference is the way in which the population is defined. In quantitative research we make the assumption that there is a true knowable state of the population. The true state of the population is represented by an actual parameter, such as '49% of the Australian population are females' or 'the average IQ of Canadian high school students is 104'. Qualitative research is not concerned with measuring quantities or counting frequencies but with the experiences and meanings of these experiences for individuals and communities. Sampling in qualitative research is referred to as purposive; the researchers deliberately select the participants who are best placed to provide the information for understanding the personal meanings of health-related events. For example, what is it like to survive a heart attack? What does it mean for patients to undergo transplantation surgery? Our sample provides the information which enables us to understand the emotional processes of coping with a heart attack or the psychological demands of being an organ recipient.

Let us consider an example of this approach. Say that you are intending to conduct qualitative research for clarifying issues relevant to the problem of family or 'domestic' violence. The first step would be to clarify the specific aspect of the problem you were studying and define the issues in which you were interested. For example, say that you wanted to collect data concerning the impact of physical violence on the lives of women. A key issue in defining the population is the cultural and historical setting in which it is enmeshed. The moral, legal and health-related problems involved in domestic violence are varied across cultures and change with time. In Western societies it is only relatively recently that violence in domestic settings has been recognized as a serious crime and this influences the experiences of the people involved in the events. Therefore, it is essential to define carefully the cultural characteristics of the group being studied and to keep in mind that we must be very cautious in generalizing the findings from one study to other settings and people.

Purposive sampling

Qualitative sampling is designed to be purposive, i.e. to select cases for inclusion in the research

Table 3.3 Nine of the commonly used purposive sampling methods

Sampling method	Brief description of sampling strategy
Extreme or deviant case sampling	The cases in these methods are chosen because of their deviance and the hope that this deviance may illustrate issues about more regular or mainstream cases
Maximum variation sampling	Purposefully picking a wide range of variation on dimensions of interest. Documents unique or diverse variations that have emerged in adapting to different conditions. Identifies important common patterns that cut across variations
Homogenous sampling	Focuses, reduces variation, simplifies analysis, facilitates group interviewing
Stratified purposeful sampling	Illustrates characteristics of particular subgroups previously found to be of interest. Facilitates comparisons of issues across groups
Snowball or chain sampling	Identifies cases of interest from contacts who know of suitable interview participants; that is, people who are likely to be good sources of information
Theory-based or operational construct sampling	Finding manifestations of theoretical construct of interest so as to elaborate and examine the construct
Opportunistic sampling	Following new leads during fieldwork, taking advantage of the unexpected, flexibility for discovering new issues
Random purposive sampling	(Still small sample size.) Adds credibility to sample when potential purposeful sample is larger than one can handle. Reduces judgment within a purposeful category. Unlike qualitative research, this strategy is not used for generalizations or representativeness
Convenience sampling	Similarly to quantitative sampling, this strategy saves time, money and effort. Poorest rationale, lowest credibility for findings. Yields information-poor cases

that are likely to be illustrative of particular issues or circumstances. Patton (1990) has proposed a typology of sampling that is widely used in qualitative research. Table 3.3 shows nine of the commonly used purposive sampling methods. Interested readers can follow up in Patton (1990) or Liamputtong Rice & Ezzy (1999) for more detailed information.

To illustrate qualitative sampling we have selected four strategies. Considering the example of physical violence against women in domestic settings, what would be the researcher's purpose for selecting the following strategies?

- *Extreme or deviant case sampling*. We would select respondents who experienced severe, even life-threatening, physical violence. Alternatively, we could select women who experienced minimal violence (such as verbal abuse). Both extremes would illustrate the consequences of violence from the victims' perspective and the understanding of these issues could contribute to developing sensitive and effective interventions.
- *Maximum variation sampling*. Here we would select across all the categories and severities of domestic violence which have been previously identified. Our purpose here is to identify the common themes emerging from the victims' experiences. This approach would also help to understand how different types of violence might result in varying outcomes from the individual's perspective.
- *Snowball or chain sampling*. The assumption here is that people sharing similar important experiences might become aware of each other in a community. In particular, key informants who are sympathetic to the researcher's work might be helpful in identifying 'information-rich' cases. For example, people who are actively involved in victim support groups would be a good source for further contacts with possible participants. In addition, such key informants can act as 'gatekeepers', protecting participants from unsafe

researchers. However, 'snowball' sampling can lead to the biased selection of participants and an inaccurate representation of the issues.

- *Theory-based sampling*. A researcher might hold the theory that the occurrence and experience of family violence might be shaped by patriarchal traditions. Here the researcher might purposely select couples who have more-or-less equal relationships and couples where there has been an imbalance of power. The theoretical construct of interest might be 'power' or 'oppression' and this construct might be manifested in the ways in which women experience and interpret violence.

To summarize, we want to make sure the people we are interviewing or the situations we are observing are representative of the targeted issues. We want to gain as complete a picture of the issues from the perspectives of our participants as possible. The usual approach to qualitative sampling is to interview or observe people in situations that fit our criteria until we have reached a point where little or no new observations are being made. This is called saturation.

External validity and sampling

The term *external validity* refers to the extent to which the results of an investigation can be generalized to other samples or situations. External validity can be classified into two types: population and ecological (Huck et al 1974).

Population validity

This refers to generalizing the findings from the sample to the population from which it was drawn. We have already examined the importance of having a representative sample in the generalizing of results from a sample to a population. However, an investigator in the health sciences might face another problem; that the accessible population from which the sample was taken might not be the same as the target population, that is, the one of general interest.

Let us illustrate this point with an example. A physiotherapist working in a large private maternity hospital intends to examine the effectiveness of a new antenatal exercise procedure for pregnant women for controlling levels of pain during delivery. A random sample of 50 pregnant women is chosen for the investigation from the population attending the hospital. The sample is then randomly assigned into two groups: one receiving the new antenatal exercise procedure, the other receiving the traditional programme. The researcher finds a statistically and clinically significant difference between the two procedures, such that the new programme is shown to be effective.

Strictly speaking, these findings can be generalized only to the population of women who attend the hospital. If the target population is *all* women who are having babies, then the generalization will lack external validity, because women attending different hospitals or having children at home had no chance of being included in the study sample. They might have different characteristics and these different characteristics may interact with the treatment in different ways to that of the sample. For instance, women who chose to deliver at home might respond better than those who chose to go to hospital but we would not know this from this study.

Ecological validity

There is another facet of external validity: the situation in which an investigation is carried out might not be generalizable to other situations. This is called ecological validity. Consider the following examples:

1. It has been shown that for clinical pain (due to disease or injury), morphine is an excellent analgesic. However, in laboratory studies of pain induced by electric shocks, morphine had little effect on subjects' reports of pain threshold. Clearly, generalizing from laboratory to clinical settings, or vice versa, has to be done with extreme caution (Beecher 1959).
2. Coronary arteriography involves the insertion of a small-gauge catheter into the coronary arteries, and injection of a dye for X-ray visualization. It was initially reported that the mortality rates for this rather dangerous-sounding practice were only 0.1% (1 per 1000) in a first-class medical institution. However, later reports from various other institutions showed mortality rates as high as 8% (80 per 1000) (Taylor 1979). Clearly, the effectiveness of treatments or the usefulness of clinical evaluations can well depend on who implements them.

The above examples illustrate the caution necessary in generalizing findings.

Summary

Appropriate sampling strategies ensure the external validity of studies and their findings. The aim of sampling strategies is to ensure the selection of a sample that is *representative* of the population of objects, persons or events the investigator aims to study. Incidental and quota samples are chosen for convenience, but these sampling strategies do not guarantee a representative sample. Random and stratified random sampling methods ensure that all important groups and characteristics in the population have the best chance to be selected and included in the study sample. Random sampling strategies are the most desirable to obtain a representative sample, although random sampling is not always feasible or desirable in health research. Area sampling and systematic sampling are also strategies that can be used to obtain representative samples.

An adequate sample size reduces the chance of large sampling errors. The probable sampling error is inversely proportional to the square root of the sample size. It was argued that optimum sample size is not the maximum number of obtainable subjects or a constant number or proportion. Rather, it has to be estimated for a specific investigation on the basis of the parameters of the phenomenon being studied and the study circumstances. Two types of external validity were discussed: population and ecological. External validity is related to inference, which involves using evidence from a limited set of elements to formulate general propositions.

Various approaches to purposive sampling in qualitative research were reviewed and an example of their application to a hypothetical study situation was presented.

Self-assessment

Explain the meaning of the following terms:

area sample
biased sample
ecological validity
external validity
incidental sample
population
population validity
random sample
random sampling
representative sample
sample
sampling error (probable)

True or false

1. The basic idea underlying sampling is to select a representative sample from which the investigator can make inferences to the population.
2. A sample is said to be random when it is not representative of the population.
3. If a population contains 50% males and 50% females, and our sample has 10% males and 90% females, then our sample is said to be biased.
4. When you take a patient's blood pressure daily, you are in fact sampling from a population of potential blood pressure readings.
5. If the patient's blood pressure fluctuates considerably during the day, then a single reading will be an inadequate sample.
6. Stratified random sampling involves the selection of the most accessible elements of the population.
7. Say that important subgroups of the population are known and subjects are sampled incidentally in proportion to these subgroups. This sampling method is known as area sampling.
8. A random sample is one in which 50% of the elements of a population have equal chances of being sampled.
9. Random sampling in health sciences is the least expensive and time-consuming strategy for selecting a sample.
10. When you take blood pressures hourly, you are in fact using a random sampling method.
11. The larger the sample size, the larger the sampling error.
12. The problem of internal validity might emerge when we generalize results obtained in one research setting to another.
13. An incidental sample cannot be, in principle, representative of the population.

14. If a sample is representative, it yields valid data for making generalizations about the population from which it was drawn.
15. The main difference between random and quota sampling is that in quota sampling particular subgroups of the population are represented proportionately.
16. Incidental sampling generates less bias than systematic sampling.
17. Area sampling involves selecting an area at random and assessing the inhabitants of that area.
18. A sample that is unbiased is a representative sample of the population from which it is drawn.
19. Sample error decreases as the sample size increases.
20. If the sample size is halved, the sampling error will be doubled.
21. If a sample is large (say $n > 500$) then the sample must be representative.
22. Generalizing findings from laboratory to clinical settings raises questions of ecological validity.
23. The problem of population validity refers to a population which contains invalid elements.

Multiple choice

1. As sample size increases:
   a the sampling error decreases
   b the ecological validity of the investigation increases
   c the population becomes more accessible
   d the sample becomes more biased.
2. A representative sample:
   a consists of at least 500 cases
   b must be a random sample
   c is defined as the inverse of the square root of the sample size
   d reflects precisely the crucial dimensions of a population.
3. An incidental sample is:
   a not necessarily biased
   b generally obtained through costly and difficult sampling procedures
   c used only in non-experimental investigations
   d none of the above.

An investigator wishes to study individuals suffering from agoraphobia (fear of open spaces). The investigator places an advertisement in the paper asking for subjects. One hundred replies are received, of which the investigator randomly selects 30. However, only 15 subjects actually turn up for their appointment.

Questions 4–6 refer to the above information.

4. Which of the following statements is true?
   a The final 15 subjects are likely to be a representative sample of the subjects selected by the investigator.
   b The final 15 subjects are likely to be a representative sample of the population of agoraphobics.
   c The randomly selected 30 subjects are likely to be a representative sample of those agoraphobics who replied to the newspaper advertisement.
   d None of the above are true.
5. The problem with drawing a representative sample of subjects with clinical conditions such as agoraphobia is that:
   a the subjects who consent to participate may be unrepresentative of the target population
   b no sampling strategies are appropriate
   c no complete lists of sufferers' names are usually available
   d *a* and *c*.
6. The basic problem confronted by the investigator is that:
   a the accessible population might be different from the target population
   b the sample has been chosen using an unethical method
   c the sample size was too small
   d agoraphobics are impossible to study in a scientific way.
7. Say that it is known that coronary disease occurs twice as frequently among males as females and three times more commonly among over 50-year-olds than those under 50. Given a sample of 120 obtained by quota sampling, how many subjects would you expect to be female and under 50?
   a 60
   b 40
   c 30
   d 10.
8. Referring to the population in Question 7, and given a sample obtained by stratified random sampling, how many females over 50 would you expect in the sample?
   a 40
   b 30
   c 10
   d 5.

9. If a study is externally valid then:
   a its results can be generalized to other equivalent settings
   b it must have been an experiment
   c quota sampling must have been used
   d all the subjects in the sample must have been equivalent.
10. A random sample is one in which:
   a all the elements had an equal chance of selection
   b a chance method was used to select the elements included in the sample
   c both *a* and *b*
   d neither *a* nor *b*.
11. When we say that a study lacks ecological validity we are implying that:
   a the study was carried out in a laboratory
   b the results cannot be generalized to other settings
   c the target population is different to the accessible population
   d all of the above.
12. Pilot studies are useful for establishing:
   a the approximate sample size necessary for the investigation
   b the appropriateness of the sampling model being employed
   c both *a* and *b*
   d neither *a* nor *b*.
13. If a pilot study indicates that the effect is likely to be small in relation to the sampling error then the investigator should:
   a abandon the research project
   b use a relatively small sample
   c use a relatively large sample
   d use an incidental method of sampling.

Chapter Four

4

Causal research and internal validity

CHAPTER CONTENTS

Introduction

Some types of research have purely descriptive goals: the aim of the investigator is to gather data, make observations and measure phenomena. For example, we may gather statistics on the incidence or prevalence of schizophrenia or neonate deaths within a community. However, in the health sciences there is often a need to identify the causes of illness and disabilities, not just to describe them. It is by understanding these causes that we can formulate and justify our assessments and interventions. Furthermore, we can justify the interventions we use in that we can point to evidence that demonstrates that the interventions are causing the beneficial changes in patients' symptoms. In health research the concept of internal validity is related to the design of research projects and the extent to which we can unambiguously identify causal relationships between interventions and outcomes.

The specific aims of this chapter are to:

1. Examine the concept of causality.
2. Examine how threats to internal validity generate plausible alternative explanations.
3. Discuss how the use of control groups can overcome extraneous variables.

Causality

For philosophers the notion of causality is difficult and controversial. In health research we need to take a more pragmatic approach and look for the following criteria for demonstrating causal relationships so that we can determine to what extent an intervention has caused the observed outcomes. The criteria for demonstrating causality are:

1. The cause must precede (occur before) the effect.
2. The cause and effect co-vary. If the cause occurs then so does the effect.
3. If the cause does not occur, then the effect does not occur.

The first criterion is quite simple. For example, if we say that injury to the person's arm is the cause of that person's reported pain, we assume that the injury was sustained prior to the onset of the pain. Clearly, if the pain had been already present, the injury would not be seen as the cause of the pain. Second, we assume that there will be concomitant variation between the injury and the pain. The worse the injury, the more severe the pain. As the injury recovers, a decrease in the level of pain can be expected. In general, we are establishing a lawful relationship between the cause and the effect.

However, observing a relationship between two events is not sufficient to demonstrate causality. For example, night follows day in a predictable, lawful fashion, but we do not say that day causes the night. We must attempt to eliminate plausible alternative explanations or hypotheses that offer rival causal explanations for the findings. For example, pain might persist even after the injury has healed. There might be other variables operating which maintain a person's experience of pain.

Threats to internal validity in intervention studies

Cook & Campbell (1979) defined a number of threats to the internal validity of intervention studies, including experiments and quasi-experiments. These threats to internal validity compromise the ability of the researcher to conclude that the different interventions administered to groups of subjects are in fact responsible for the differences or lack of differences that are observed. Threats to internal validity are sources of alternative explanations for the outcome of an investigation.

The main threats to internal validity include:

1. *History.* This refers to events that intervene between the pre-intervention test and post-intervention test that do not form part of the treatment or intervention being investigated by the researcher. For example, in a study of the effects of an exercise programme on hypertension, some of the patients might take up additional exercise, such as playing tennis. The effects occurring as a result of these unplanned activities are history effects.
2. *Maturation.* In a study over time, the patients may naturally mature or change. This is a particular issue with paediatric and geriatric populations and studies that are conducted over an extended time period (longitudinal studies). Maturation refers to any time-dependent internal changes in respondents or patients which might influence the outcome. For example, simple infection will resolve without interventions through the action of the immune system.
3. *Testing.* The patient may, as a result of familiarity with the testing procedures, appear to improve spontaneously. These are sometimes called practice effects. For instance, the re-administration of an IQ test might lead to better performance without an actual improvement in the subjects' intellects.
4. *Instrumentation.* During the time between measurements, the measuring instrument might change, for example start reading heavier or lighter, leading to apparent improvement or deterioration when no such change has occurred.
5. *Regression to the mean.* This is a special effect that originates from the unreliability of test measures. Often, clinical research involves the selection for study of patients who have achieved particularly low or high scores on one or more measures, for example the most depressed patients or those with the highest measured cholesterol levels. It is often found that such patient groups, when retested, will show apparent spontaneous improvement (or decline) known as regression towards the mean.
6. *Selection or assignment errors.* The groups being compared may be different at the outset prior to any intervention because of inadequate assignment or selection procedures. If the groups are different to start with, then this difference is likely

to persist after the intervention. Thus a post-intervention difference may not reflect the effects of the intervention when the groups were different at the start.

7. *Mortality*. You do not have to have anyone die in a study to have research study mortality! Mortality in a study refers to when a participant withdraws from the study before its completion. There may be more dropouts in one group than in others, making the groups different. For example, participants in a placebo group might reject an ineffectual treatment and refuse to participate to the conclusion of the study. Since the participants who drop out might be different from those who stay, the experimental and the control group no longer remain equivalent. Once again, if the groups are different, it is difficult to ascribe these differences to their different treatments.

The following investigation illustrates that some of the above threats to validity generate plausible alternative explanations for outcomes. A recent study attempted to demonstrate the effects of an exercise programme in patients with occlusions (blockages) of some of the major arteries in the leg. The 'effect' or the dependent variables measured included the distance walked by the patient to 'limits to the pain tolerance'. The independent variable or the 'cause' was the exercise programme, which included daily walking, with encouragement to increase the distance daily. In addition the patients were also strongly advised to stop smoking and were given a diet low in animal fats and carbohydrates. More than half the patients were smokers and reported that their smoking declined markedly by the end of the programme. The procedure for the study involved pre-tests on the dependent variables (including walking), the administration of the treatment, and then a re-test six weeks after the programme commenced. The results (Fig. 4.1) appeared to indicate an increase in the walking distance. The study was done under the guidance of a vascular surgeon, who also selected the patients.

Figure 4.1 shows that there was an increase in the patients' average walking distance. The question here is: 'Was this change in the dependent variable (effect) caused by the independent variable or by other uncontrolled extraneous variables?' Extraneous variables are variables other than the independent variables which plausibly influence the

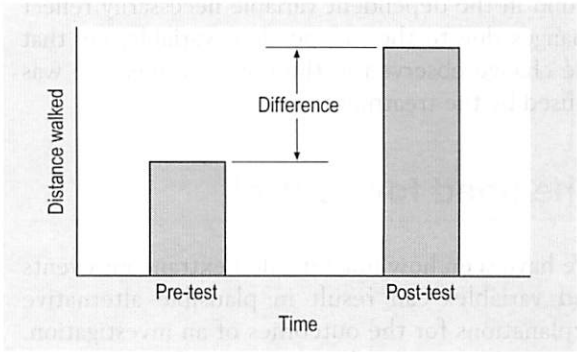


Figure 4.1 • Walking tolerance: before and after exercise programme.

dependent variable. As discussed earlier, uncontrolled extraneous variables are the sources of threats to internal validity.

1. *History*. The threat here is from simultaneous changes in walking, smoking pattern, and diet, any of which might have been responsible for some or all of the changes in the dependent variable reported in the study.
2. *Maturation*. The degree of this threat depends on the natural history of the illness. Natural improvements (or, for that matter, deterioration) in the condition might account for at least some of the difference in walking.
3. *Testing*. Both the pre-test and the exercise programme included walking. It is possible that the difference between pre-test and post-test is simply due to practice effects, the patients becoming more confident in walking to real tolerance limits.
4. *Instrumentation*. This threat is not relevant in this instance, in that no sophisticated or potentially inaccurate measuring devices were used to measure the dependent variable.
5. *Regression to the mean*. Possibly the vascular surgeon selected extreme cases for the study: those most in need of treatment. Their performances might have drifted back towards the mean on the post-test.

6–7. *Selection errors and mortality*. These threats are not relevant, as no control group was included in the hypothetical study.

We have not shown that the extraneous variables described above necessarily caused the reported difference. Rather, because of the lack of control, the investigator cannot claim that the differences

found in the dependent variable necessarily reflect changes due to the independent variable, i.e. that the change observed in the outcome measure was caused by the treatment.

The need for control

We have seen how uncontrolled extraneous events and variables can result in plausible alternative explanations for the outcomes of an investigation. In causal research, the adoption of an appropriate design potentially enables us to remove the confounding influences of extraneous variables. If we can do this, i.e. correctly attribute any effects we observe to our interventions and eliminate the effects of other uncontrolled factors upon outcome, we can say that the experiment is internally valid or has internal validity.

In the context of laboratory research, an investigator can use several types of strategy to achieve control over extraneous variables. For example, a physicist studying electricity will make sure that the apparatus is insulated against extraneous electrical disturbances; a researcher studying biofeedback would make sure that the subjects were in a noise-insulated, temperature-controlled room. In 'field' research, in applied clinical settings, such tight control is not feasible. Other methods, such as the inclusion of control groups in the design, are used to maximize internal validity.

A control group consists of subjects that undergo exactly the same conditions as the group receiving the intervention under investigation. For example, in drug trials control group participants will often receive an injection of saline solution, if the experimental treatment is administered via injection, in order to control for the effects of actual injection. If the medication were administrated orally, similar-looking inert tablets would be used for control subjects. It has been found that if people receive any form of 'therapy', improvement may occur even when the 'treatment' or intervention is physiologically and chemically inert. This is referred to as a placebo effect. The control group allows the researcher to measure the size of the placebo effect, and to take it into account when interpreting the study results.

If we are to include a control group in our intervention studies, it is essential that at the outset the experimental and control groups are as similar as possible. We have to take the group of subjects participating in the study and split them up into the experimental and control groups as equally as possible. This process is called assignment. The assignment of subjects to their groups by the investigator is an essential feature of an intervention study.

The use of control groups in clinical research

Let us re-examine the investigation outlined on page 45 in terms of the impact upon internal validity when we use a control group in the study.

1. Assess the walking distances of the patients in a pre-treatment test.
2. Assign subjects into experimental or control groups by matching on the basis of pre-test performance, and random assignment within pairs.
3. Administer:
   (a) experimental intervention (total package of exercise plan, walking, diet and smoking reduction) and
   (b) control intervention (an alternative activity, walking, diet and smoking reduction). The *only* difference between the two groups is that one group receives the exercise programme, and the other some alternative activity.
4. Test both groups on walking distance, following the treatment.

The results of this fictional study are presented in Figure 4.2.

Let us examine how this new design stands up to threats of internal validity in contrast to the original investigation outlined on page 45.

1. *History.* Both groups had walking, smoking decrement and diet; it is unlikely that the difference between post-test results shown in Figure 4.2 is due to these variables.
2. *Maturation.* Both control and experimental groups had the same time to recover or deteriorate; it is unlikely that this factor explains the difference.

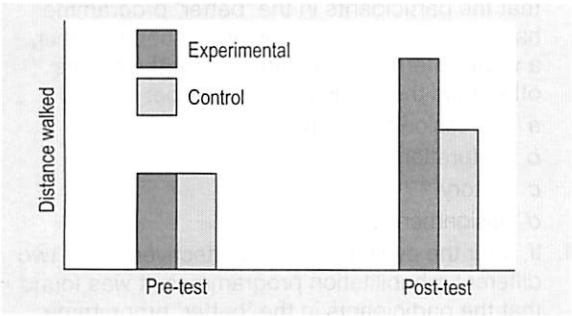


Figure 4.2 • Hypothetical results of a pre-test/post-test study.

3. *Testing*. Both groups had the pre-test and the walking; it is unlikely that this factor explains the difference.
4. *Instrumentation*. Not relevant to this study.
5. *Regression to the mean*. Both groups were similar on pre-test performance, given that they were matched.
6. *Selection error*. Random or matched assignment would have controlled for this, so initial differences between the two groups is minimized.
7. *Mortality*. This depends on the actual data; internal validity is preserved if drop-out rates are equivalent in the two groups.

We can see that using a design with a control group resulted in an improvement of the internal validity of the investigation. Note that the use of a control group has not removed the effects of history and other extraneous variables in this study. It is still possible, but much less likely, that the differences in the outcomes for the two groups were not a result of the different interventions they received. Now the investigator has a much sounder basis for deciding whether or not the exercise programme was effective.

It should be noted that there are ways, other than the use of control groups, to minimize the effects of uncontrolled extraneous variables. For example, if 'noise' is a possible extraneous variable in administering a test, we may 'insulate' from it by using a quiet setting. However, if noisy settings are the norm in real life, then the search for higher internal validity may be at the cost of external validity.

Summary

The demonstration of causality involves the discounting of plausible alternative explanations for the study outcomes. If there are uncontrolled extraneous variables, these may generate plausible alternative explanations to the research hypothesis that the intervention alone was responsible for the study outcomes. In this case the investigation is said to have low internal validity. When designing a study to investigate the causal effects of an intervention we aim to reduce the threats to internal validity. When the design includes an appropriate control group, the ability to attribute causal effects to the intervention and not to other extraneous events becomes more convincing.

However, in field research involving human subjects, complete control over the phenomena is not possible. To some extent, control might have to be sacrificed in order to ensure the external validity of the results. In the next chapter, we examine a variety of research strategies and the designs used within these strategies aimed at maximizing internal and external validity.

Self-assessment

Explain the meaning of the following terms:

assignment errors
causality
causal explanations
control
dependent variables
extraneous variables
history
independent variables
instrumentation
internal validity
maturation
mortality
regression to the mean
testing

True or false

1. It is possible to have an effect occur prior to its cause.

2. If a group of children were found to have improved their reading skills in a long-term treatment programme, it would be easy to eliminate maturation as a rival explanation to treatment for their improvement.
3. If two groups in a study start off on an unequal basis, this would probably be an example of assignment errors.
4. In a test of skill that is easily learned, testing effects are unlikely to be a problem.
5. Control groups will eliminate assignment errors.
6. If the researcher can attribute the outcomes in a study to the treatment programmes employed and not other factors, then the study is internally valid.
7. If a study is internally valid then it must be externally valid.
8. Mortality refers to the difficulty of patients dying while undergoing the treatment programme.
9. Regression to the mean occurs mainly in studies where groups have been selected to participate on the basis of some form of extreme score in a pre-test.
10. Control refers to the need to eliminate alternative conflicting explanations for the outcomes observed in a study.

Multiple choice

1. If a group of participants in a study are selected on the basis of a particularly poor pre-assessment, it is likely that they will appear to 'improve' when re-assessed soon after, without any treatment. This phenomenon is known as:
   a regression to the mean
   b maturation
   c history
   d assignment error.
2. If, after a 12-month programme, a group of children with reading problems had improved their reading ages by an average of 9 months, a viable alternative explanation for the results, other than the effectiveness of the programme, would be:
   a regression to the mean
   b maturation
   c history
   d assignment error.
3. If, after the evaluation of the effectiveness of two different rehabilitation programmes, it was found that the participants in the 'better' programme had all attended a disability workshop together, a viable alternative explanation for the results other than the programme would be:
   a regression to the mean
   b maturation
   c history
   d assignment error.
4. If, after the evaluation of the effectiveness of two different rehabilitation programmes, it was found that the participants in the 'better' programme were selected from a less-impaired group, then a viable alternative explanation for the results other than the programme would be:
   a regression to the mean
   b maturation
   c history
   d assignment error.
5. The advantage of including a control or comparison group in a study of treatment effectiveness is that:
   a effects such as maturation can be eliminated
   b effects such as maturation can be reduced
   c effects such as maturation can be measured
   d effects such as maturation can be ignored.
6. If a well-designed study demonstrates a convincing advantage of one therapeutic technique over another, but is based on a sample of five people in the two groups, then the study is likely to have:
   a high internal and low external validity
   b high external and low internal validity
   c low internal and external validity
   d high external and internal validity.
7. Those patients that receive inert treatment in a control comparison group yet respond as if they had received real treatment are demonstrating:
   a the placebo effect
   b external validity
   c internal validity
   d regression to the mean.
8. In a long-term treatment study with a single group with no comparison or control group, it is possible to attribute any improvements observed to:
   a the treatment
   b history effects
   c maturation
   d all of the above.

Section Two

Discussion, questions and answers

Our next discussion question takes us back over 50 years in time, to the setting of the closed psychiatric institutions where persons with serious mental illnesses were often confined under dismal and overcrowded conditions. Psychiatric researchers were desperately seeking new and effective treatments which could enable the residents to return to the community and thus to relieve the pressure on the institutions. A team of Italian researchers, led by the psychiatrist Cerletti, were working on a new technique which they hoped would provide a quick and effective treatment for schizophrenia.

Unfortunately, their research programme was guided by what was later shown to be a false hypothesis: that persons who suffered from epilepsy did not develop schizophrenia. On the basis of this false hypothesis it was predicted that inducing epileptic convulsions would help to reduce the signs and symptoms of schizophrenia. Thus, in the late 1930s thousands of mentally ill persons were administered convulsants, such as the drugs cardiazol and metrazol. The drugs induced convulsions that were difficult to control and proved to be very dangerous. Cerletti argued that the use of electrical shocks to induce epileptic convulsions would be a safer approach than the drugs.

In an article which outlines the historical development of electroshock, Krzyzowski (1989) reported what happened when Cerletti and his assistant Bini first presented these ideas to their colleagues.

> *Bini, at a conference in Munich in 1936, and Cerletti in the same year in Milan, mention the possible application of electric current to cause therapeutically desired epileptic attacks. The idea was almost unanimously rejected on the grounds of its barbarity and associated hazard. It should be noted that the electric chair had just been introduced in America.*
>
> (Krzyzowski 1989, p. 51)

Despite this hostile reaction, Cerletti and his team continued research into electric shock, experimenting with animals in Rome's slaughterhouses. According to Krzyzowski (1989, pp 51–52) the team was ready for their first human subject by 1938.

> *On April 15 1938, a patient manifesting distinct symptoms of mental illness was admitted into the clinic in Rome after having been arrested by the police for travelling on a train without a ticket. The condition of the patient was then as follows: fully normal orientation, expressed distinct introversion and persecution delusions often using neologisms. He considered himself to be under telepathic influence directing his behaviour. At the same time he exhibited hallucinations thematically related to the delusions. He was depressed*

and altered neurological conditions were found. Schizophrenia was diagnosed.

Having selected their first subject, the team administered the first electroconvulsive therapy (ECT) procedure.

Two electrodes were attached symmetrically in the vicinity of the crown and forehead and then a relatively low 80V current was passed for 1.5s. On switching the current on, the patient sat upright on the bed, his muscles contracted and he fell back onto the bed, not losing consciousness, however. He cried out for a while and then became quiet.

It appeared that the voltage applied was too low to induce the convulsion required for the therapeutic effectiveness of the ECT procedure. What happened next is discussed by Krzyzowski:

Continuation of the treatment was postponed until the next day. The patient, on hearing of such suggestions, roused himself and shouted normally: 'No more. It could kill me'. The words were spoken aloud normally, while previously he used only a specific and hardly understandable jargon, self-devised and full of neologisms. This normal utterance confirmed and assured Cerletti of the effectiveness of the method and, in spite of the strong reservations of his assistants, he decided to repeat the ECT without delay. This time a 110V current with a pulse duration of 1.5s was used. Once again a short lived, general contraction of all muscles was observed followed by a full classic epileptic type attack of convulsions. All those present uneasily watched the pallidness and cyanosis accompanying the attack, relaxing as the patient gradually recovered.

The story, you will be pleased to hear, had a happy ending.

After prolonged treatment consisting of 11 full and 3 incomplete shocks, the patient was discharged from the clinic in good health. At follow-up a year later, the mental state of the patient was seen to be good and stable. Subsequent years evidenced a widening application of ECT.

(Krzyzowski 1989, p. 52)

As a matter of interest, ECT is still used by contemporary psychiatrists, but in a greatly modified fashion:

- It is used with anaesthetized patients, with electrodes placed only on one side of the head.
- It is not used with persons with schizophrenia (for whom it was found to be ineffective) but rather with persons who are profoundly depressed.
- It is used as a 'last resort' when current pharmaceutical treatments are ineffective.

Questions

The following questions concerning research planning are related to the above narrative:

1. On the basis of the information given propose aims and/or hypotheses which might have guided the above study.
2. Given the state of psychiatric knowledge and practice of the late 1930s, do you think the above study was justified?
3. Give three or four reasons why a contemporary ethics committee might reject such a research project.
4. By present ethical standards, what should have been done after the patient shouted 'No more, it could kill me'?
5. Cerletti proceeded with a more severe shock on the grounds that the previous shock 'improved' the patient's condition. Comment on this logic, in the context of scientific methodology.
6. Do you think such dangerous experiments are ever justified in the context of health care? If yes, under what conditions?
7. Discuss two or three problems with the 'internal validity' (Ch. 4) of this study. Suggest simple changes in 'control' which may help to improve internal validity.
8. Comment on the 'external validity' of this study in the light of the fact that in subsequent studies ECT was shown to be ineffectual as a treatment for schizophrenia.
9. Do you think Cerletti and his colleagues were guided by a rather simplistic paradigm of schizophrenia? Explain, by comparing the biomedical and biopsychosocial approaches discussed in Section 1.

Answers

1. The research seems to have been guided by several interrelated aims and/or hypotheses.
   (a) The first aim was to find the electrical shock intensity that was sufficient to induce an epileptic seizure in a human.
   (b) The second aim was to demonstrate that the seizure did not result in death or disability; that is, that the treatment was 'safe'.
   (c) The hypothesis implicit in the research might be stated as: 'A course of ECT is effective for reducing the signs and symptoms of schizophrenia'.

   You may be able to suggest other aims and hypotheses. None of the outcomes was stated precisely in the paper quoted.
2. As you may have judged from this brief excerpt, the state of knowledge concerning the biological causes of mental illness was confused and had a weak empirical basis. Biological treatments, such as drug-induced epileptic seizures, were poorly theorized, dangerous and generally ineffective. In this context, in the late 1930s, it could be argued that experimenting with new and safer treatments was justified.
3. We outlined some relevant ethical guidelines in Chapter 2 in terms of which the present study would be judged as problematic, e.g. questionable benefits for patients, lack of informed consent, dangerous and painful intervention and lack of consultation with the relatives/guardians of a mentally ill person.
4. Obviously, discontinue the research. Even though the person was confused when admitted to the hospital, his request was rational and reasonable. There was no doubt whatsoever concerning the patient's desires, and by present ethical standards researchers must comply with such requests.
5. The first electrical shock did not induce the epileptic convulsion which was postulated as the factor 'causing' the therapeutic change. As the first shock simply hurt the man, there was no theoretical or empirical justification for his apparent improvement being due to this shock.
6. One could argue that risking people's health and lives in the context of 'heroic' medicine is never justified. Rather, we should look to prevention or gentler, more natural treatments. The other point of view is that aggressive medical treatment is justified if there are incapacitating and chronic problems, such as schizophrenia. According to this approach, painful and potentially harmful experimental treatments are justified provided that the study is well designed and the participants are well informed and have consented.
7. The issue is whether or not the apparent improvement in the patient's condition was due to the ECT or to other 'extraneous' factors or variables. There are several possibilities which provide plausible alternative explanations, such as:
   (a) *History*. The patient may have been frightened by the treatment and 'pretended' to be better to escape the situation.
   (b) *Maturation*. The patient may have recovered anyway: his condition may be cyclical.
   (c) *Testing or instrumentation*. There may have been inaccuracies and bias in the way in which the patient's condition was assessed.

   Control may be introduced by using groups of persons who have (a) no treatment and/or (b) another treatment for schizophrenia. The appropriate designs for showing causal effects are outlined in Section 3. However, because of the poorly designed research in the area, there were almost two decades of useless treatments before it became evident that ECT was not an effective treatment for schizophrenia.
8. Clearly, there was no evidence that the improvement claimed in this study was caused by ECT. In addition, it is tricky to make inferences from unrepresentative samples to populations. In this study, the patient may have shown symptoms of depression, which perhaps responded to the treatment. But this may not be generalizable to persons who show other patterns of schizophrenia. External validity is ensured by appropriate sampling procedures (Ch. 3) and clear operational definition and assessment of the condition, as outlined in Chapter 12. Without appropriate sampling and assessment procedures, we may use inappropriate treatments, unsuitable for the specific needs of our patients.
9. The researchers were working in the context of a 'biomedical' model, assuming that schizophrenia was simply a biological disorder which could be suddenly cured by a heroic treatment such as ECT. A biopsychosocial approach takes a more complex view of chronic disorders, and research programmes include identifying and treating both psychological disabilities and social handicaps entailed in schizophrenia. Biopsychosocial research may involve both quantitative and qualitative methods.

Section Three

Research designs

SECTION CONTENTS

In the previous section, we identified broad categories of descriptive (non-experimental) and causal (experimental) strategies for conducting research. The aim of Section 3 is to discuss these issues in more detail by examining various research designs and outlining their applications for conducting health sciences research. As in other areas of creative human activity, such as architecture or the fashion industry, designs refer to explicit plans for completing an object or an activity.

The conceptual basis for experiments is the need for control, as outlined in Chapter 4. In experimental research (see Ch. 5) we manipulate one or more variables (independent variables), while controlling the extraneous variables, by using appropriate controls such as control groups. If the experiment is well designed and properly conducted we are in a position to demonstrate the causal effects of the independent variables on the outcome or dependent variables. The random assignment of subjects to experimental groups is a common way to achieve control. However, health and illness outcomes often have multiple and interacting causes requiring multivariate experimental designs.

Another issue we will address is reactivity of human beings to social situations, such as being involved in research. Appropriate experiments attempt to control for the biases of both the participants and the experimenters involved in a research project.

There are situations where for practical or ethical reasons we cannot randomly assign subjects to control or treatment groups. Here we use quasi-experimental methods which involve comparing pre-existing groups undergoing different treatments (see Ch. 6). There are other research designs discussed in Chapter 6, including naturalistic comparisons, correlational designs and surveys. Naturalistic comparisons and correlational designs enable us to describe and predict relationships among variables, but should be used only with extreme caution in attempting to understand causal effects. Surveys that include the use

of questionnaires and interviews to study a person's knowledge, attitudes and beliefs concerning aspects of health and illness are essentially a tool for descriptive research. The accuracy of the information collected within the framework of non-experimental designs depends to a large extent on the use of appropriate sampling strategies for selecting the partipants.

An interesting and commonly used design in clinical settings is the $n = 1$ or 'single case' design. The advantage is that using $n = 1$ designs we may be able to demonstrate causal effects using only one or two subjects, without the need for separate control groups. The major limitation of $n = 1$ (and other types of clinical case studies) is that the findings may not be generalizable to other cases or situations.

Chapter 8 describes some of the elementary characteristics of qualitative field research designs. This chapter aims to compare and contrast quantitative and qualitative research, and describes the importance of qualitative research for understanding the personal experiences of our clients and/or patients. Qualitative data collection strategies (such as in-depth interviewing and participant observation) are outlined in Section 4 and quantitative data analysis in Section 6.

Chapter Five

5

Experimental designs and intervention studies

CHAPTER CONTENTS

Introduction

Experiments are an important form of intervention studies. A well-designed experiment enables researchers to demonstrate how manipulating one set of variables (*independent* variables) produces systematic changes in another set of variables (outcome or *dependent* variables). Experimental designs attempt to ensure that all variables other than the independent variable(s) are controlled; that is, that there are no uncontrolled extraneous variables that might systematically influence or bias changes in the outcome variable(s). Control is most readily exercised in the sheltered environment of research laboratories and this is one of the reasons why these settings are the preferred habitats of experimentalists. Much of clinical experimentation, however, takes place in field settings (e.g. hospitals and clinics), where the phenomena of health, illness and treatment are naturally located. Even in these natural settings researchers can exercise control over extraneous variables. The randomized controlled trial (RCT) is an experimental research design aimed at assessing the effectiveness of a clinical intervention. The experimental approach has also been used more broadly in the 'hard' sciences.

The specific aims of this chapter are to:

1. Examine the basic structure of experimental research designs and RCTs.
2. Consider threats to the validity of the results obtained from experiments and RCTs.

Experimental research

Experimental studies and RCTs involve the following steps:

1. *Definition of the population.* Researchers define the population to which they wish to generalize. For example, this might be males over 55 years with coronary heart disease or the local community or a certain type of health care organization.
2. *Selection of the sample.* Using an appropriate sampling method, the study sample is selected from the population. It is desirable that the sample is representative of the population. It is important to note that steps 1 and 2 are common to all research designs.
3. *Assignment procedures.* Using a randomized assignment procedure the participants are allocated to the intervention groups. Random assignment means that each participant has equal chance of being assigned to any of the groups. In an RCT, one of the groups receives a control intervention with the active component of the intervention missing. For example, in an experimental study of the effectiveness of different types of weight-loss programmes, one group might receive an instruction manual, another might receive supervised dietary training and a third group may receive no intervention. The purpose of the randomized assignment procedure is to ensure that the groups are as similar as possible on all the relevant variables. If the groups are substantially different, it will be very difficult to attribute any differences in final outcome to the 'treatments' being administered.
4. *Administration of intervention (treatment).* The researcher then administers the intervention(s) to the various groups participating in the experiment. This is called the independent variable(s). It is important that the intervention is administered in an unbiased way, in order that a fair test of any differences in outcomes may be provided. As will be discussed later, awareness of the expected outcomes on the part of participants may lead to a spurious biasing of the outcomes for those participants. Therefore, the true aims and expectations of the researcher are sometimes concealed from the participants in an attempt to control expectancy effects.
5. *Measurement of outcomes.* The researcher assesses the outcome of the experiment through measurement of the dependent or outcome variable(s). Sometimes, the dependent variable is measured both before and after the experimental intervention (this is called a pre-test/post-test design) and other times only afterwards (this is called a post-test only design).

Thus, in an experimental study the researcher actively manipulates the independent variable(s) and monitors outcomes through measurement of the outcome or dependent variable(s).

Assignment of participants into groups

Random assignment

The simplest approach is to assign the participants randomly to independent groups. Each intervention group represents a 'level' of the independent variable. Say, for instance, we were interested in the effects of a new drug (we will call this drug A) in helping to relieve the symptoms of depression. We also decide to have a placebo control group, which involves giving patients a capsule identical to A, but not containing the active ingredient.

Given a sample size of 20, we would assign each participant randomly to either the experimental (drug A) or the control (placebo control) group. Random assignment could involve encoding all the 20 names into a computer program and using random numbers to generate two groups of 10. In this case, we would end up with the two equivalent groups:

	Levels of the independent variables	
	Control group (Placebo)	Experimental group (Drug A)
Number of participants	$n_c = 10$	$n_e = 10$

Here n_c and n_e refer to the number of participants in each of the groups, given that the total sample size (n) was 20. We can have more than two groups if we want. For example, if we also included another drug 'B', the independent variables would have three levels. We would require a total of 30 participants if the group sizes remained at 10.

Matched groups

Random assignment does not guarantee that the two groups will be equivalent. Rather, the argument is that there is no reason that the groups should be different. While this is true in the long run, with small sample sizes chance differences among the groups may distort the results of an experiment or RCT. Matched assignment of the participants into groups minimizes group differences caused by chance variation.

Using the hypothetical example discussed previously, say that the researcher required that the two groups should be equivalent at the start of the study on the measure of depression used in the experiment. Using a matched-groups design the participants would be assessed for level of depression before the treatment and paired for scores from highest to lowest. Subsequently, the two participants in each pair would be randomly assigned to either the experimental group or the placebo control group. In this way, it would be likely that the two groups would have similar average pre-test depression scores.

Different experimental designs

Four types of experimental design will be discussed. These are the post-test only, pre-test/post-test, repeated measures and factorial designs.

Post-test only design

At first it may appear that this would make measurement of change impossible. At an individual level this is certainly true. However, if we assume that the control and experimental groups were initially identical and that no change had occurred in the controls, direct comparison of the post-test scores will indicate the extent of the change. This type of design is fraught with danger in clinical research and should only be used in special circumstances, such as when pre-test measures are impossible or unethical to carry out. The assumptions of initial equivalence and of no change in the control group often may not be supported and, in such cases, interpretation of group differences is difficult and ambiguous.

Pre-test/post-test design

In this design, measurements of the outcome or dependent variables are taken both before and after intervention. This allows the measurement of actual changes for individual cases. However, the measurement process itself may produce change, thereby introducing difficulties in the clear attribution of change to the intervention on its own. For example, in an experimental study of weight loss, simply administering a questionnaire concerning dietary habits may lead to changes in those habits by encouraging people to scrutinize their own behaviour and hence modify it. Alternatively, in measures of skill, there may be a practice effect such that, without any intervention, the performance on a second test will improve. In order to overcome these difficulties, many researchers turn to the post-test only design.

Repeated measures

In order to economize with the number of participants required in an experimental design, the researchers will sometimes re-use participants in the design. Thus, at different times the participants may receive, say, drug A or drug B. If it were the case that every subject encountered more than one level of the drug variable or factor, then 'drug' would be termed a repeated measures factor. An important consideration is using a 'counterbalanced' design to avoid series effects. For example, half the participants should receive drug A first, and half drug B. If all the participants received drug A first and then drug B, the study would not be counterbalanced and we would not be able to determine whether the order of administration of the drugs was important. Time is a common

Table 5.1 Examples of a factorial design

	Drug A	No drug
Rogerian therapy	1	2
No psychological therapy	3	4

repeated measures factor in many studies. A pre-test/post-test design involves the measurement of the same participants twice. If 'time' is included in the analysis of the study, then this is a repeated measures factor. In statistical analysis, repeated measures factors are treated differently from factors where each level is represented by a separate, independent group. This is true both for matched groups, discussed earlier, as well as for repeated measures, discussed above.

Factorial designs

A researcher will often not be content with the manipulation of one intervention factor or variable in isolation. For example, a clinical psychologist may wish to investigate the effectiveness of both the type of psychological therapy and the use of drug therapy for a group of patients. Let us assume that the psychologist was interested in the effects of therapy versus no psychological treatment, and of drug A versus no treatment. These two variables lead to four possible combinations of treatment (see Table 5.1).

This design enables us to investigate the separate and combined effects of both independent variables upon the outcome measure(s). In other words, we are looking for interactions among the two (or more) factors. If all possible combinations of the values or levels of the independent variables are included in the study, the design is said to be fully crossed. If one or more is left out, it is said to be incomplete. With an incomplete factorial design, there are some difficulties in determining the complete joint and separate effects of the independent variables.

In order that the terminology in experimental designs is clear, it is instructive to consider the way in which research methodologists would typically describe the example design in Table 5.1. This is a study with two independent variables (sometimes called factors), namely, type of psychological therapy and drug treatment. Each independent variable or factor has two levels or values. In the case of psychological therapy, the two levels are Rogerian therapy and no psychological therapy. This would commonly be described as a 2 by 2 design (each factor having two levels). There are four groups (2×2) in the design.

If a third level, drug B, was added to the drug factor, then it would become a 2 by 3 design with six groups required. Two groups, drug B with Rogerian therapy and drug B with no psychological therapy, would be added. It is possible to overdo the number of factors and levels in experimental studies. A $4 \times 4 \times 4$ design would require 64 groups: that is a lot of research participants to find for a study. It is relevant to note that when we evaluate two or more groups over a period of time we are also using a factorial design.

Multiple dependent variables

Just as it is possible to have multiple independent variables in an experimental study, it is also sometimes appropriate to have multiple outcome or dependent measures. For example, in order to assess the effectiveness of an intervention such as icing of an injury, factors such as extent of oedema and area of bruising are both important outcome measures. In this instance, there would be two outcome measures. The use of multiple dependent variables is very common in health research. The outcomes measured are usually evaluated individually, although there are more complex statistical techniques which enable the simultaneous analysis of multiple dependent variables.

External validity of experiments and RCTs

We have already discussed external validity in Chapter 3. There are further criteria for ensuring the generalizability of an investigation, depending on the procedures used and the interaction between the patients and the therapist.

The Rosenthal effect

A series of classic experiments by Rosenthal (1976) and other researchers have shown the importance of expectancy effects, where the expectations of the experimenter are conveyed to the experimental subject. This type of expectancy effect has been termed the *Rosenthal effect* and is best explained by consideration of some of the original literature in this area.

Rosenthal and his colleagues performed an experiment involving the training of two groups of rats in a maze-learning task. A bright strain and a dull strain of rats, specially bred for the purpose, were trained by undergraduate student experimenters to negotiate the maze. After a suitable training interval, the relative performances of the two groups were compared. Not surprisingly, the 'bright' strain significantly outperformed the 'dull' strain.

What was surprising is that the two groups of rats were actually genetically identical. The researchers had deceived the student experimenters for the purposes of the study. The students' expectations of the rats likely had resulted in different handling and recording which had apparently affected the rats' measured learning outcomes. Rosenthal's results have been confirmed time and time again in a variety of experimental settings, and with a variety of participants.

If the Rosenthal effect is so pervasive, how can we control for its effects? One method of control is to ensure that the 'experimenters' do not know the true purpose of the study; that is, the experimental hypothesis. This can be done by withholding information – just not telling people what you are doing – or by deception.

Deception is riskier and less ethically acceptable. Most organizations engaged in research activity have ethics committees that carefully monitor and limit the use of deception in research. If the people carrying out the data collection are unaware of the research aims being tested we say that they were 'blind' to the research aims.

The Hawthorne effect

As well as the impact of the expectations of participants in experimental studies, there is also the issue of whether attention paid to participants in the experimental setting alters the results.

In the late 1920s, a group of researchers at the Western Electric Hawthorne Works in Chicago investigated the effects of lighting, heating and other physical conditions upon the productivity of production workers. Much to the surprise of the researchers, the productivity of the workers kept improving independently of the actual physical conditions. Even with dim lighting, productivity reached new highs. It was obvious that the improvements observed were not due to the manipulations of the independent variables themselves, but some other aspect of the research situation. The researchers concluded that there was a facilitation effect caused by the presence of the researchers. This type of effect has been labelled the *Hawthorne effect* and has been found to be prevalent in many settings. Of particular interest to us is the Hawthorne effect in clinical research settings. It must be considered that even 'inert' or useless treatments might result in significant improvements in the patients' condition under certain circumstances. The existence of the placebo effect reinforces the importance of having adequate controls in applied clinical research. Although we cannot eliminate it, we can at least measure the size of it through observation of the control group, and evaluate the experimental results accordingly.

Controlled research involving human participants

Investigations involving human participants require that researchers should consider both psychological and ethical issues when designing experiments. Human beings respond actively to being studied. When recruited as a research participant, a person might formulate a set of beliefs about the aims of the study and will have expectations about the outcomes. In health research, placebo effects are positive changes in signs and symptoms in people who believe that they are being offered effective treatment by health professionals. These improvements in signs and symptoms are in fact elicited by inert treatments probably mediated by the patients' expectations. Health professionals

who hold strong beliefs about the effectiveness of their treatments communicate this attitude to their patients and possibly increase the placebo effects.

Placebo responses as such are not a problem in everyday health care. Skilled therapists utilize this effect to the patients' benefit. Of course, charlatans exploit this phenomenon, masking the poor efficacy of their interventions. In health research, however, we need to demonstrate that an intervention has therapeutic effects greater than a placebo, hence the need for a placebo control group. The ideal standard for an experiment is a *double-blind* design, in which neither the research participants nor the health providers/experimenters know which participants are receiving the active form of the treatment.

It is not always possible, however, to form placebo control groups. In the area of drug research and some other physical treatments it can be possible; but with certain behaviourally-based interventions, in areas such as psychotherapy, physiotherapy or occupational therapy, double-blind placebo interventions may be impossible to implement. For instance, how could a physiotherapist offering a complex and intensive exercise programme be 'blind' to the treatment programme being offered? In these situations researchers employ no treatment or traditional intervention control groups when evaluating the safety and effectiveness of a novel intervention.

Also, there are ethical issues that need to be taken into account when using placebo or non-intervention control groups. Where people are suffering from acute or life-threatening conditions, assignment into a placebo or no-treatment control group could have serious consequences. This is particularly true for illnesses where prolonged participation in a placebo control group could have irreversible consequences for the sufferers. Under such circumstances, placebo or no-treatment controls might well be unethical and the selection of a traditional treatment group would be preferred.

Summary

The experimental (RCT) approach to research design involves the active manipulation of the independent variable(s) through the administration of an intervention often with a non-intervention control group and the measurement of outcome through the dependent variable(s). Good experimental design requires careful sampling, assignment and measurement procedures to maximize both internal and external validity.

The Hawthorne and Rosenthal effects are important factors affecting the validity of experimental studies, and we attempt to control for these by 'blinding' when ethically possible.

Common experimental designs include the pre-test/post-test, post-test only, repeated measures and factorial approaches. These designs ensure that investigations can show causal effects. More recent approaches to evidence-based health care have emphasized the importance of employing experimental trials wherever possible.

Self-assessment

Explain the meaning of the following terms:

control group
double blind
factorial design
Hawthorne effect
history
independent variable
instrumentation
internal validity
matching
mathematical model
maturation
mortality
physical model
placebo effects
post-test only design
pre-test/post-test design
random assignment
Rosenthal effect
selection error

True or false

1. The dependent variable is the variable measured by the investigator.

2. Selection or assignment errors arise when, after the assignment of the participants, the groups are not equivalent.
3. If in a clinical investigation more people die in the control group than the experimental group, the investigation lacks internal validity.
4. Ideally, the control group and the experimental group should receive exactly the same treatments.
5. Given that a placebo is an inert substance, its administration has no effects on the participants' behaviours.
6. The random assignment of participants is always preferable to assignment by matching.
7. Persons can serve as their own controls.
8. We reduce the effects of subject and experimenter expectancies by blindfolding.
9. Dependent variables are manipulated by the researcher.
10. Control can be exercised by keeping the values of extraneous variables constant.
11. Confounding extraneous variables generate explanations other than the independent variable.
12. Placebo effects can be seen as confounding extraneous variables.
13. If A antecedes B, then A must be the cause of B.
14. 'Double-blind' experiments are impossible for evaluating some health interventions.
15. 'Maturation', that is natural recovery, can be an extraneous variable.
16. The use of placebo control groups is always unethical.

Multiple choice

1. The aim of controlled observation is to:
   a remove the effects of confounding influences
   b identify the effects of the independent variable on the dependent variable
   c establish causal relationships
   d all of the above.
2. To say that an investigation lacks internal validity means that:
   a the independent variable had no effect
   b the dependent variable was not measured
   c uncontrolled variables may have affected the outcome
   d there were several dependent variables.
3. Which of the following is most representative of a placebo effect?
   a Headache is reduced when an antidepressant is administered.
   b Headache is reduced one second after swallowing an analgesic, well before it is absorbed by the body.
   c Headache is increased after a fierce argument with a 'significant other'.
   d Headache is decreased after the use of biofeedback.

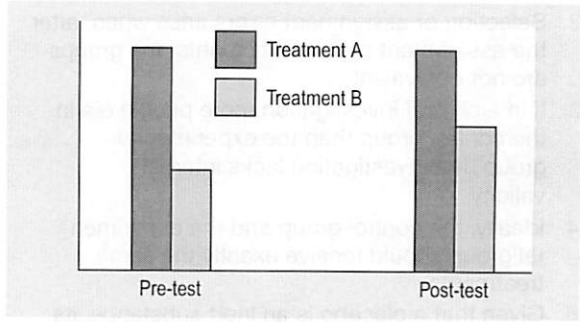
A researcher is studying the effect of a new drug on healing of ulcers. Patients are assigned randomly, by the physician who treats them, to receive either the standard treatment or the new drug. Patients are informed that they are being studied, but they do not know which treatment they are getting. The measure of rate of healing is the number of days until the ulcer is completely healed.

4. The independent variable in the above study was:
   a ulcers
   b the new drug
   c type of treatment
   d rate of healing.
5. The dependent variable in the above study was:
   a ulcers
   b the new drug
   c type of treatment
   d rate of healing.
6. This study is:
   a double blind because the patients do not know what treatment they are getting or which is expected to be more effective
   b double blind because neither the researcher nor the patients know what treatment each person is getting
   c single blind because the patients do not know what treatment they are getting but the physician who treats them does
   d not blind at all as the patients know they are being studied.
7. Which of the following threats to internal validity is not controlled for in this study?
   a maturation
   b regression to the mean
   c repeated testing
   d history
   e none of the above.

8. This investigation is an example of:
   a an experiment, because patients are assigned randomly
   b a study, because the treating doctor did the assigning
   c an experiment, because two treatments are compared
   d a study, because there is no control group.
9. If an experiment is internally valid, this means that:
   a the study's results cannot be generalized to other equivalent settings
   b the sampling method is appropriate
   c the study is not really an experiment
   d the independent variable is responsible for any trends observed
   e the dependent variable has face validity.

A researcher wishes to study the effectiveness of a new drug for treating arthritis. Participants are selected randomly from patients attending a clinic and are assigned randomly to treatment with the new drug (A) or with the current standard treatment (B). A pre-test/post-test design is used. A number of arthritic symptoms are assessed using a checklist of seven items.

10. The dependent variable in this study is:
    a the number of arthritic symptoms
    b the new drug
    c the method of assignment
    d the type of treatment.
11. The independent variable in this study is:
    a the number of arthritic symptoms
    b the new drug
    c the method of assessment
    d the type of treatment.
12. The actual outcome of the study is illustrated by the following graph. Even though the participants were assigned randomly, the graph indicates possible threat(s) to internal validity due to:
    a maturation
    b regression to the mean
    c assignment
    d *b* and *c*
    e all of the above.
13. A factorial design involves:
    a more than one independent variable
    b more than one dependent variable



    c only one independent variable, but with more than one level
    d no control group
    e none of the above.
14. Simulation research means:
    a using computers to analyse data
    b studying differences between different artificial limbs
    c studying a model of a situation to reach conclusions about a situation
    d pretending to research one behaviour while actually studying another.
15. An appropriate design for an investigation will be one which:
    a minimizes all possible sources of error
    b is experimental
    c gives you the answer you expect
    d none of the above.
16. If a study is externally valid, then:
    a it must have been an experiment
    b the dependent variable has face validity
    c it cannot be internally valid
    d its results can be generalized to other equivalent settings.
17. The placebo effect:
    a happens only in drug studies
    b occurs in experimental as well as control groups
    c is another name for relaxing participants
    d occurs only if you do not have double blinding.
18. Random selection of participants in a study is typically employed to:
    a maximize generality of the results
    b minimize random measurement error
    c control assignment errors
    d minimize the Rosenthal effect
    e minimize the Hawthorne effect.

19. Random assignment of participants in an experiment is typically employed to:
    a minimize random measurement error
    b minimize the Rosenthal effect
    c minimize the Hawthorne effect
    d ensure that the experimental and control groups are similar at the outset
    e maximize generality of the results.

20. The principal advantage of a factorial design is that:
    a all important factors are taken into account
    b only one factor is considered
    c the dependent variable is more reliable
    d the joint effects of two or more independent variables may be assessed
    e the independent variable has more levels.

Chapter Six

6

Surveys and quasi-experimental designs

CHAPTER CONTENTS

Introduction

In the previous chapter, the basic features of the experiment were presented. Experiments or randomized controlled trials (RCTs), when properly designed and implemented, allow the study of causal relationships between variables. However, there are many other legitimate research designs. Although the RCT is held up by some commentators, such as those using the Cochrane Collaboration framework, as producing the highest level of research evidence, most health sciences research does not use an experimental design. For example, Bloch (1987) reported that of the 757 articles on back pain interventions published in 1985, only eight employed RCT or experimental designs. In this chapter we will examine a variety of approaches used for investigation of research questions that do not lend themselves to an experimental approach.

The specific aims of this chapter are to:

1. Examine the uses of naturalistic non-experimental designs.
2. Discuss the use of surveys in health sciences research.
3. Outline the characteristics of designs involving naturalistic comparisons, correlations and quasi-experiments.
4. Identify some of the limitations of these designs.

Naturalistic non-experimental research designs

If experimental designs are supposed to provide the tightest possible control of extraneous factors, why should we resort to alternative non-experimental research methods? There are a number of reasons why non-experimental research designs ought to be employed instead of experimental designs.

1. Many variables are not amenable to experimental manipulation, i.e. they are static. For example, if the research question is concerned with sex differences in responses to heart surgery, then sex cannot be manipulated by the researcher. Similarly, if the researcher is interested in age differences, the ages of the participants cannot be altered by the researcher.
2. Often, it is ethically inappropriate to investigate research questions using an experimental design. For example, if a researcher wished to perform a study on the effects of smoking upon health, studying this in an experimental design would require the experimenter to allocate participants randomly to the smoking or non-smoking group. Clearly, it is unethical to force some participants to smoke and others not to smoke. In experimental designs using a non-treatment control group, valuable and effective treatment might be withheld from participants. This situation involves serious ethical concerns that might lead to an experiment not being approved.
3. Experiments are best used to study simple causal relationships between variables, i.e. does the intervention cause positive therapeutic effects? However, many human diseases and illnesses are not determined by a single cause but rather by a number of causes interacting in a complex fashion. For example, heart disease may be caused by factors such as smoking, excessive stress, inappropriate diet or genetic factors. To identify such possible causal (or risk) factors, we need to study systems as they function in nature. That is, we should investigate patients in their natural setting, even with the difficulties this entails.

Naturalistic non-experimental designs are appropriate alternatives to experimental methods. In the health sciences, experimental methods are focused on a rather narrow band of research questions that relate to the effectiveness of interventions. There are many other interesting research questions that do not relate to intervention effectiveness and, even when the questions do relate to interventions, there are circumstances in which non-experimental methods are more appropriate.

Surveys

Surveys are investigations aimed at describing accurately the characteristics of populations for specific variables (see Fig. 2.1). Surveys are commonly used in health research for the following purposes:

1. To establish the attitudes, opinions or beliefs of persons concerning health-related issues. The data collection techniques often include questionnaires or interviews.
2. To study characteristics of populations on health-related variables, such as utilization of health care, blood pressure, emotional problems or drug use patterns.
3. To collect information about the demographic characteristics (age, sex, income, etc.) of populations. A government census can be an important source of knowledge concerning population characteristics.

The statistics obtained from surveys can present us with an overview of the patterns of states of health, illness and the use of health services in a given community. In this way, we can gain insights into issues such as the prevalent causes of death or the health-related requirements of the population. The outcomes of the surveys can be the bases for hypotheses and theories concerning the causes of illness in a community. The area of health science concerned with such matters is called *epidemiology*.

Epidemiology

Epidemiology is the field of study that is concerned with the distribution and determinants of health and illnesses in groups of people. Last (1988) defined epidemiology as:

> *The study of the distribution and determinants of health related states or events in specified populations and the application of this study to control of health problems.*

Epidemiology is a frequently used approach in public health as it is concerned with the study of the health of populations. It differs from much clinical research in that it is oriented to the population or group rather than individual level.

Epidemiology goes back to the time of Hippocrates who was concerned with the effects of environments upon the health of populations. The work of John Snow in the 19th century concerning the epidemiology of cholera is also considered to be a major landmark in the development of the discipline. Snow mapped the distribution of cholera cases in London and demonstrated that the water supply to different houses supplied by different water companies was involved with the transmission of the disease. The existence of the cholera organism was inferred from these observational data.

Epidemiology focuses on specific target populations and compares the occurrence of different health or disease states within these populations. Thus the numbers of cases, new cases and the relative risk of having the target condition are key components of the epidemiological approach. Two key terms in epidemiology are incidence and prevalence. The *incidence* of a disease is the rate at which new cases occur in a specified population during a specified period. The *prevalence* of a disease is the proportion of a specified population that have the target condition at a specified point in time.

These basic definitions are often used to construct indicators of population health that can be compared across different countries. The World Health Organization (WHO) maintains statistics concerning rates such as the birth rate, the infant mortality rate and perinatal mortality rate. The WHO web-sites provide a wealth of comparative epidemiological data (http://www.who.int/en/).

Epidemiology is also concerned with how diseases are transmitted or spread within populations. The classic epidemiological approach conceptualizes disease within the framework of host, agent, vector and environment. The host is the human, the agent is the infection or health problem, the vector is the means by which the disease is carried, e.g. in the case of malaria this may be a mosquito, and the environment is the setting or mechanism that promotes the exposure. Disease can be transmitted directly (from person to person) or indirectly (via water, such as was demonstrated in Snow's classic study of London water supplies). Although this framework was developed initially within the context of infectious diseases, it is actually much more broadly applicable.

The pattern of outbreaks of disease within populations is also an area of interest to epidemiologists. Common patterns of occurrence of diseases include endemic, epidemic and pandemic patterns. A disease is said to be endemic to a particular area if it habitually occurs within the area. An epidemic is the occurrence of a disease that is over and beyond the usually expected rate. We often talk of flu epidemics, meaning an unexpectedly high rate of the occurrence of influenza. A pandemic is a worldwide epidemic. We have become acutely aware of pandemics such as human immunodeficiency virus/acquired immuno-deficiency syndrome (HIV/AIDS) and other feared pandemics such as severe acute respiratory syndrome (SARS), and, more recently, avian flu.

Epidemiology uses both experimental and non-experimental designs for collecting evidence. Beaglehole et al (2000) provide a useful discussion of equivalences of epidemiological and other research terminology.

Naturalistic comparison studies

Essentially, this type of study involves the comparison of naturally occurring groups or populations. For example, a researcher may be interested in the relative performances of males and females on a test of spatial abilities. The researcher might take a sample of each sex studying at a university and then compare the relative performances of the males and females on the test. Alternatively, a researcher might be interested in whether people growing up in different cultures have different pain reactions. Here, the researcher would select a sample of volunteers from the cultural groups of interest and compare their pain responses to standard pain stimuli.

There are extraneous variables which can be controlled in this type of investigation. In these

studies, the researcher can control variables such as the ages or educational backgrounds of the participants by selecting similar groups. However, the researchers have not actually changed or manipulated the variables being studied. All that has been done is the measure of differences between naturally occurring groups.

To relate this to our first example, if our researcher determines that there is indeed a statistically significant difference in the performance of males and females on the measure(s) chosen, well and good. If our researcher then claims that biological factors are solely responsible for the differences, this is another matter altogether. This inference is only one of a number of alternative explanations that could be advanced to account for the differences observed. It is possible, for instance, that males and females in a sample have been exposed to different types of toys and games during their development, so that males might perform better or worse on some tasks. In the second example, any significant results in the pain responses might not be due to 'culture', but rather to systematic biological differences among the groups or a complicated interaction between these factors.

The simple fact of the matter is that it is difficult to demonstrate causation in natural comparison studies. However, if consistent evidence emerges from a logical sequence of studies, the investigators will gain crucial information about the differences between groups on clinically relevant measures. Natural comparison studies are vital components of the researcher's methodological tools.

Correlational studies

The aim of correlational studies in the health sciences is to identify interrelationships among variables. In epidemiology these studies are referred to as 'ecological' designs. A correlation is a statistic that expresses numerically the strength of association that exists between two or more variables.

Let us look at a simple illustration of correlational or epidemiological studies. Say that clinical observations indicate that people who suffer from coronary heart disease tend to be overweight. Such observations might generate the hypothesis: 'There is a positive correlation between being overweight and the probability of coronary heart disease'. Here, the investigator will need to draw a representative sample of the population of interest (let's say 500 men and 500 women randomly selected from a population of healthy men and women aged 40 living in a specified district). The next step would be to measure the participants' weights and heights. These measures might be monitored over a period of time to check for drastic changes in weight. The second variable would be measured by the criterion of whether or not the participant suffered from heart disease during a specified period (e.g. ten years), representing the length of the study. The incidence of coronary disease can then be converted into a probability for a particular category of weight (see Table 6.1).

It can be seen from Table 6.1 that the higher an individual scores on one variable (percentage overweight), the higher the scores on the other variable (probability of coronary heart disease). In this way, the fictional data presented in Table 6.1 are consistent with the predictions of the hypothesis stated above.

Two points should be considered at this stage. First, no evidence has been presented that one variable is causally related to another. This type of investigation does not, by itself, allow us to conclude that, for instance, 'being overweight causes coronary heart disease'. There are several alternative hypotheses, such as 'stress causes both being overweight and heart disease', which can also account for the findings. Second, it is clear that, at least for the period of the investigation, the variable 'percentage overweight' does not account for all of the other variable 'probability of coronary heart disease'. That is, according to the data presented in Table 6.1, there are normal or underweight individuals who suffer from coronary heart disease and very overweight people who do not. Clearly, there must be other variables which influence the incidence of coronary heart disease. There are many other variables such as smoking, blood cholesterol level, personality type, stress and family history of heart disease that also correlate with the probability of coronary heart disease.

Some diseases, for example lung cancer and chronic back pain, have complex, interacting causes. Natural comparison, correlational and

Table 6.1 Fictional data representing the relationship between variables 'percentage overweight' and 'probability of coronary heart disease' (for a 10-year period)

Percentage overweight	Number of participants	Number suffering coronary heart disease	Probability of coronary heart disease
Underweight or normal	600	30	30/600 = 0.05
10–19% overweight	200	20	20/200 = 0.10
20–29% overweight	100	15	15/100 = 0.15
30–39% overweight	75	20	20/75 = 0.27
40% or more overweight	25	10	10/25 = 0.40

ecological designs can help us to identify risk factors; that is, factors that might elevate the risk of the onset or progress of these illnesses.

Quasi-experimental designs

If preventive interventions are to be undertaken to reduce the risk factors associated with a disease, then we require reasonable evidence that these factors are, in fact, causally related to the disease. *Quasi-experimental* designs are often used for this purpose.

Quasi-experimental designs can resemble experiments, with the important difference that there is no random assignment into treatment groups. However, the investigator can control the time at which a treatment is introduced or withdrawn. One such method is a time-series design. This type of design is used by both clinical and epidemiological researchers.

Time-series designs

Time-series designs involve repeated observations before and after a given treatment. In this way, changes in the sequence of observations following the introduction of a treatment may represent the effects of the treatment on the observed variable. Let us look at an example illustrating the use of time-series designs in health sciences research.

Returning to the risk factor of 'being overweight' we discussed previously, the following investigation using a time-series design could provide evidence for a causal relationship between this variable and cardiac disease.

1. Select an appropriate population to study.
2. Specify the dependent variable; that is, some clear-cut measure of 'coronary heart disease'. A commonly used measure may be the incidence of the disease. By 'incidence' we mean the number of new cases of the disease reported in relation to the population within a specified period of time (for example, 50 per 100 000 per year).
3. Introduce an appropriate treatment which reduces the magnitude of the risk factor. In our example, a health promotion package could be introduced, emphasizing exercise and good eating habits. Let us assume that this intervention is adequately financed and a significant proportion of the population adheres to the programme. It could then be hypothesized that introduction of the health promotion package will result in a decrease in the incidence of coronary heart disease in the community.
4. Monitor the dependent variable over a period of time. It is essential to have readings of the variable both before and after the introduction of treatment. In this instance, the incidence of coronary illness would be determined from the medical records of hospitals, clinics and physicians. Public health authorities often gather and make available such statistics.

Epidemiologists are involved in the monitoring and surveillance of the health of human populations to track the occurrence of existing and possibly new health problems. The existence of HIV/AIDS was established very early by epidemiologists at the

US Centers for Disease Control using sophisticated monitoring and surveillance systems. At the heart of such systems are sophisticated data-gathering procedures.

While the language and concepts of epidemiology are different from the classical clinical disciplines, the approaches to data collection are quite similar to those disciplines. Epidemiologists rely upon the collection of (especially longitudinal) survey data as the basis for their analyses of the distribution and transmission of diseases, are involved in the analysis of RCTs and, like clinicians, attempt to build causal and statistical models of disease occurrence.

Figure 6.1A and B represents two of the many possible empirical outcomes using time-series designs. We will assume that the incidence of the illness was monitored for six years before and after the introduction of the treatment.

There are, of course, other possible outcomes. However, in this case, Figure 6.1A would be consistent with the predictions of our hypothesis, while the outcome shown in Figure 6.1B would be inconsistent with our hypothesis. Discussion of the way in which data generated by time-series designs are analysed is beyond the scope of the present text. Briefly, it involves the analysis of trends (increase or decrease) found in the dependent or outcome variable.

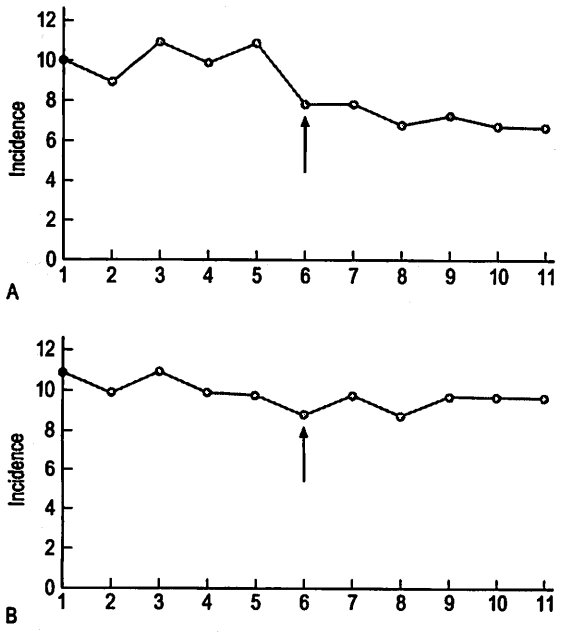


Figure 6.1 A,B • Possible outcomes of time-series designs.

Time-series designs may have problems with internal validity. What assurance have we that the decrease in the incidence of coronary heart disease, shown in Figure 6.1A, was caused by the introduction of the health promotion programme? Perhaps there was another cause, such as the introduction of drugs to control high blood pressure.

Multiple-group time-series designs

The introduction of multiple-group time-series designs involves the comparison of two or more naturally occurring groups. Let us examine a simple illustration of such designs by a fictional further investigation of the coronary heart disease problem.

Using a multiple-group time-series design, the investigator would select a community which is as similar as possible on demographic variables (socio-economic classes, age, size of the community, etc.) to the community being studied. In this way we would have:

- *Community A.* Control. Do not introduce health promotion programme.
- *Community B.* Introduce health promotion programme.

Figure 6.2 shows two of the several possible outcomes for such an investigation. Some researchers would call this type of design a prospective cohort design; prospective refers to the fact that the data are collected after the commencement of the intervention and cohort refers to the fact that we have multiple intervention/non-intervention groups in the study.

Where the two communities show different trends, the outcome shown in Figure 6.2A is consistent with our previously stated hypothesis. However, in Figure 6.2B there is a trend for decrease in both communities A and B, therefore the evidence is not consistent with the prediction of the hypothesis.

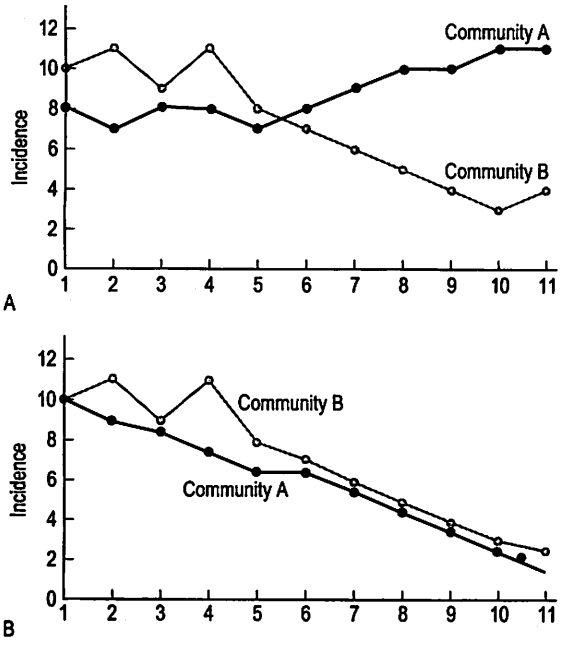


Figure 6.2 A,B • Possible outcomes of multiple-group time-series designs.

Even with multiple-group time-series (prospective cohort) designs, some threats to internal validity remain. There are no guarantees that the communities A and B were equivalent on all relevant factors, or that there were no important changes during the study in the communities which might have influenced the incidence of coronary heart disease. The best that researchers can do is to identify and try to estimate the effect of such extraneous factors on the dependent variable. There are problems such as insulating the two groups: when people in community B learn about the programme being carried out in community A they, by themselves, might initiate aspects of the programme.

The internal and external validity of naturalistic designs

We have noted several types of problems concerning the internal validity of investigations using natural comparison, correlational and time-series designs. These are similar to the problems of internal validity found in experimental RCT investigations. However, because researchers have generally less control over the phenomena being investigated, the use of natural comparison designs makes it more difficult to evaluate causal hypotheses. That is, given uncontrolled extraneous variables, a variety of plausible alternative hypotheses might be offered to account for the findings. Therefore, in areas such as epidemiology, researchers use evidence arising from a variety of investigations using different types of designs to evaluate their theories and models of the causes of human diseases.

When evaluating evidence from natural comparison designs, the external validity of the findings must be considered. In Chapters 3 and 5, we stated that external validity refers to the generalizability of the results of an investigation. Strictly speaking, the results of an investigation should be generalized only to the population from which the sample was selected. For instance, consider the example discussed previously, where a sample of male and female students were compared on a test of spatial ability. Any differences between these two groups can be generalized validly only to the population of students from which the samples were drawn. Investigators sometimes forget this obvious point, and try to make inferences about males and females in general. Such sweeping generalizations are invalid. For instance, other cultures with alternative child-rearing practices might well have males and females with completely different relative spatial abilities. Just because a variable is found to be a risk factor in one community does not guarantee that it will have the same influence on diseases in another community. Clearly, the finding that cigarette smoking is a serious risk factor for coronary heart disease in Western societies does not necessarily mean that cigarette smoking is a risk factor for coronary heart disease among other cultural groups and settings. Given the complex, interacting causes of coronary heart disease, there may be different risk factors in communities which follow different lifestyles or have different physical constitutions. It is not surprising, therefore, that advances in knowledge concerning

systematic differences between culturally, racially or sexually different groups have been slow and controversial.

Summary

Descriptive surveys were shown to be important ways of studying the attitudes, behaviours and health problems of a community. Evidence of this kind can be used to propose theories or test hypotheses of the nature and causes of illness. We examined several types of non-experimental research designs including epidemiological approaches, naturalistic comparisons, correlational and ecological designs, and time-series and prospective cohort designs. It was argued that these designs are appropriate alternatives to experimental designs, especially in field settings. However, as with experimental RCT designs, investigations using these designs must confront the same problems of internal and external validity.

Self-assessment

Explain the meaning of the following terms:

cohort studies
correlational studies
ecological designs
epidemiology
incidence
multiple-group time-series
naturalistic comparisons
prevalence
quasi-experimental designs
risk factors
surveys
time-series

True or false

1. If it is impracticable to manipulate an independent variable, the investigator might adopt a non-experimental design.
2. The finding that the incidence of birth abnormality has increased in a community since the fluoridation of the water supply is sufficient evidence that fluoridation caused the birth defects.
3. If there is a high correlation between levels of stress and heart disease, we can say that stress is a risk factor in heart disease.
4. Time-series designs involve the repeated observation of participants before and after treatment.
5. An important difference between experimental and quasi-experimental designs is that in a quasi-experimental study the participants are not assigned into treatment groups by the investigator.
6. A quasi-experimental design affords greater control over extraneous variables than an experiment.
7. It might be invalid to infer that there are general differences between males and females on the basis of findings involving a sample.
8. An investigator compares the IQs of two ethnic groups, but does not control for systematic differences in socioeconomic status. Therefore, the study lacks internal validity.
9. When we speak of the 'insulation' of one group from another, we mean attempting to prevent participants in a control group learning about and adopting the treatment given to an experimental group.
10. If a risk factor is, in fact, a cause for a disorder, we can predict that reducing the risk factor should lead to a reduction in the incidence of the disorder.
11. Surveys are utilized in the social, but not in the medical, sciences.
12. 'Naturalistic' designs are useful when it is physically or ethically impossible to manipulate the independent variable.
13. The advantage of naturalistic designs over experimental designs is that they enable the investigator to exercise more control over the independent variable.
14. Naturalistic designs should not be employed to investigate conditions which have complex, interacting causes.
15. Naturalistic comparisons yield data which unequivocally establish the causal effects of the independent variable.
16. A risk factor is a possible cause of a disease.
17. Quasi-experimental designs can involve the production of 'time-series'.
18. Time-series designs involve the calculation of correlation between time and space.

19. A multiple-group time-series design involves the comparison of two or more naturally occurring groups over a period of time.
20. A multiple-group time-series design can be employed only if the participants can be randomly assigned into treatment groups.
21. For a factor to be recognized as a 'risk factor' it must be shown that it correlates with aspects of a disorder across cultures.
22. Epidemiologists are interested in the frequency and causes of diseases.

Multiple choice

1. Naturalistic designs are appropriate when:
   a it is ethically inappropriate to manipulate the independent variable
   b only a small sample of participants is available
   c the independent variable is extremely difficult or impossible to manipulate
   d *a* and *c*
   e all of the above.
2. Naturalistic comparisons differ from experimental designs in that:
   a random selection of participants is not possible with naturalistic comparisons
   b random assignment of participants is not necessary with naturalistic comparisons
   c naturalistic comparisons lack external validity
   d all of the above.
3. A risk factor is:
   a a sufficient cause for a complex disorder
   b a necessary cause for a complex disorder
   c both a and b
   d neither *a* nor *b*.

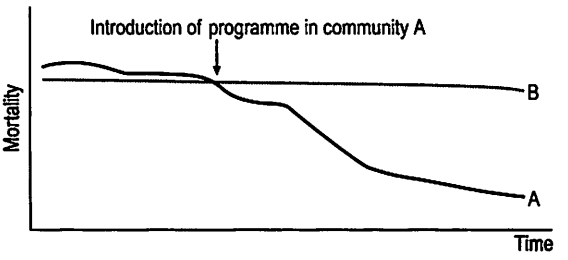
We compare a representative sample of Kamchatkans and Patagonians currently resident in Australia on a test of visual acuity.

Questions 4 and 5 refer to the above.

4. The design employed in the above investigation is:
   a a naturalistic comparison
   b a true experiment
   c a factorial design
   d a time-series design.
5. Given that the sample of Kamchatkans are found to have higher visual acuity than the sample of Patagonians, we can conclude that:
   a Patagonians have poor eyesight
   b Kamchatkans have higher visual acuity than Patagonians
   c living in Kamchatka improves one's visual acuity
   d Kamchatkans living in Australia probably have higher visual acuity than Patagonians living in Australia
   e both *a* and *d*.
6. We employ correlational designs in order to:
   a establish the magnitude and direction of association between variables
   b establish unequivocally causal effects
   c identify the possible causes of disease entities
   d both *a* and *c*.
7. We intend to investigate if the use of ice is more effective than an exercise programme in treating a particular type of physical injury. The design most appropriate for investigating this problem is:
   a an experimental design
   b a correlational design
   c a model
   d a time-series design.
8. An investigator intends to establish if there is a relationship between levels of atmospheric lead and learning deficits in children. The design most appropriate for investigating this problem is:
   a an experimental design
   b a correlational design
   c a model
   d a time-series design.

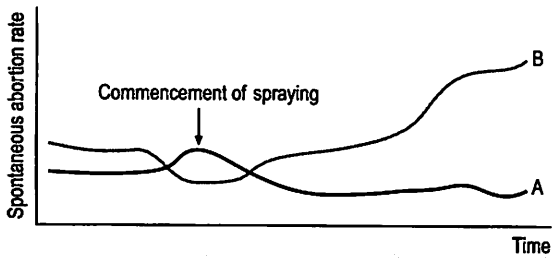
The graph below summarizes changes in infant mortality rates over a period of years in two communities, A and B. In community A, a programme of intensive fetal monitoring was introduced at the time indicated by the arrow. No such programme was introduced in community B.

Questions 9 and 10 refer to the above information.



9. The investigation should be described as:
   a a correlational study
   b an experiment
   c a time-series design
   d a multiple-group time-series design.
10. The findings summarized in the above graph appear to be consistent with the conclusion that:
   a the introduction of fetal monitoring had no effect on infant mortality rate
   b the introduction of fetal monitoring decreased mortality rate
   c the introduction of fetal monitoring increased infant mortality rate
   d none of the above conclusions is consistent with the data.
11. Risk factors of diseases:
   a can be identified by correlational designs
   b are not necessarily the sole causes of diseases
   c are not necessarily the same across cultures
   d all of the above.

A multiple-group time-series design is introduced to study the effects of an insecticide on the number of spontaneous abortions. The insecticide was introduced (not by the investigator) in community A, but not in a similar community B. The graph below represents the hypothetical data.



12. On the basis of the graph above, one can conclude that:
   a there is evidence that the introduction of the insecticide increased spontaneous abortions in community A
   b there is no evidence that the introduction of the insecticide increased spontaneous abortions in community A
   c there appears to be a contamination of community B by the insecticide
   d *a* and *c*.
13. An investigator distributes a questionnaire to a sample of ex-nurses in an attempt to discover their reasons for leaving the profession. This is best described as:
   a a quasi-experiment
   b an experiment
   c a survey
   d a social model.
14. In an attempt to demonstrate the causal relationship between intake of animal fat and heart disease, an investigator studies the eating habits and heart disease rates of demographically defined groups. Which of the following (hypothetical) findings is most inconsistent with the hypothesis that 'The intake of high levels of animal fats causes heart disease'.
   a Vegetarians have a lower rate of heart disease than meat eaters.
   b Canadian males have a lower rate of heart disease than Canadian females.
   c The increase of protein and fat intake in China has correlated with an increased incidence of heart disease.
   d Eskimos, whose diet consists mainly of animal fats, have lower incidence of heart disease than Egyptians who eat mainly grains.

Chapter Seven

7

Single case (n = 1) designs

CHAPTER CONTENTS

Introduction

We have discussed research designs involving the comparison of groups of participants selected from a population. These designs provide evidence concerning the general causes of diseases, or the overall effectiveness of interventions. However, health professionals often work with individual patients and need to understand the specific causes of their problems and the effectiveness of treatments as applied to them as individuals. n = 1 designs illustrate the close relationship between the principles for conducting research and everyday clinical practice.

The aim of the present chapter is to examine single subject (n = 1) designs, as applied by a variety of health professionals in natural clinical settings. You will be able to recognize close similarities between these designs and the quasi-experimental designs discussed in the previous chapter.

The aims of this chapter are to:

1. Identify methodological similarities between clinical problem solving and designing n = 1 studies.
2. Examine the use and comparative advantages of AB, ABAB, and multiple-baseline designs.
3. Comment on the validity of n = 1 designs.
4. Demonstrate how the interpretation of single case studies is related to generally applied methodological principles.

AB designs

Let us consider a simple example to illustrate the basic procedures involved in using $n = 1$ designs. Imagine that a patient is admitted to your ward suffering from a condition that involves having a high temperature. Before an appropriate intervention is devised, the patient's temperature is recorded every 15 minutes, for 2 hours. Following this time interval, the patient is given medication to reduce temperature. The question here is: 'How do we show that the medication was effective for reducing the patient's temperature?' Obviously, we need to show that the patient's temperature had fallen following the administration of the medication. Figure 7.1 illustrates a possible outcome.

Let us assume that the drug is known to act quickly, say in 20 minutes. The evidence shown in Figure 7.1 would be clearly consistent with the hypothesis that the medication caused a decrease in the patient's temperature. Let us generalize this example to $n = 1$ designs used in various settings. Figure 7.2 illustrates the general conventions used in $n = 1$ designs.

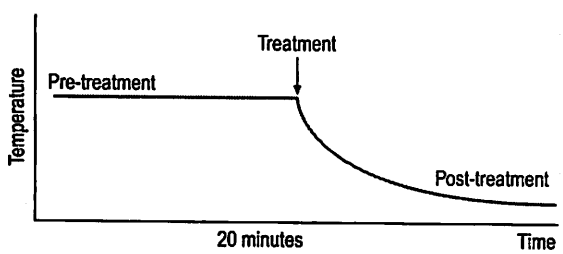


Figure 7.1 • Possible outcome of an AB design.

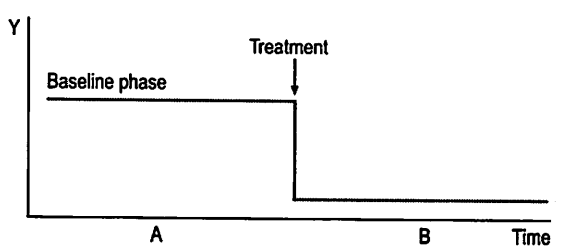


Figure 7.2 • General structure of $n = 1$ designs.

1. We can see that the y-axis (Y) represents the outcome or *dependent variable* (DV): observations made of specific physical characteristics or behaviours.
2. The x-axis represents the time over which the observations were carried out.
3. The period A represents the sessions during which no intervention was administered. This is the 'control' level of the *independent variable* (IV). The observations recorded during A form the baseline.
4. Period B represents the sessions during which intervention was administered, that is, the active stage.
5. The observations taken during A are compared with those taken during B. Systematic changes in the DV between A and B phases (increases or decreases) are assumed to reflect the influence of the intervention.

Therefore, an AB design involves taking observations during phase A, introducing an appropriate intervention, and then taking observations during B. It might have occurred to you that several of the threats to validity can be identified in AB designs. An obvious threat to validity is maturation, that is, recovery or deterioration occurring in the patient that might influence the readings on the outcome variable. Another possible threat is history, that is, influences on the patient other than the actual intervention. In the example we just looked at, one could also argue that perhaps the patient's temperature would have gone down even without the drug, because of the condition improving by itself or the environment of the ward (maturation), or perhaps the ward was air-conditioned and the patient would have cooled down anyway, drugs or not (history).

Next, we look at ABAB designs, which provide stronger control for extraneous variables than AB designs.

ABAB designs

The basic feature of ABAB designs is the alternation of intervention with no-intervention or baseline phases. That is, the researcher introduces the intervention following a baseline or no-intervention phase, then the intervention is withdrawn and then later re-introduced. Observations are recorded

during each phase and this approach permits control for the previously discussed threats to validity.

Figure 7.3 illustrates the outcomes for a hypothetical drug study using an ABAB design. When the drug is withdrawn (second A) the patient's temperature returns to previous levels. When the drug is re-introduced (second B), the patient's temperature declines. Clearly, such an outcome is consistent with a causal relationship between the independent variable (intervention) and the outcome or dependent variable (observations of temperature). Figure 7.4 demonstrates the idealized results expected using ABAB designs with a highly effective rapid-onset intervention.

Although the above design is useful for demonstrating causal effects in a single individual, there are situations where it is inappropriate. ABAB designs are not particularly useful when there is a good reason to expect that the effects for an intervention are known to be irreversible following the intervention. For example, if the medication used in the previous example involved antibiotics, then the discontinuation of such drugs after a period might not result in the re-emergence of the symptoms since the antibiotics might well cure the underlying problem. Even when reversal is possible, it might not be ethical. Clearly, if we have succeeded in establishing desirable effects in our client during the first B period, we might well be reluctant to reverse this for the sake of demonstrating causal relationships.

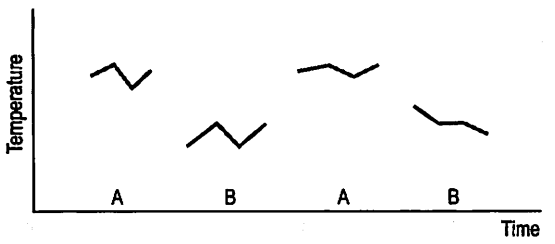


Figure 7.3 • Graph of patient's temperature under different treatment conditions.

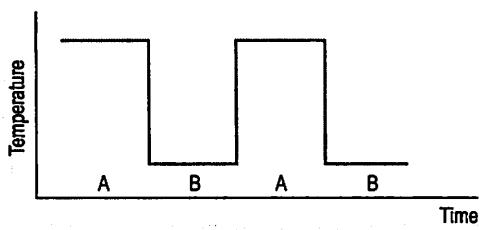


Figure 7.4 • General structure of ABAB designs.

Multiple baseline designs

Multiple baseline designs involve the use of concurrent observations to generate two or more baselines. Given two or more baselines, the investigator has the opportunity to introduce intervention affecting only one set of observations, while using the other(s) as a control. We will examine a hypothetical clinical problem to illustrate these designs.

Imagine that we have a brain-damaged client showing aggressive behaviours that disrupt therapy. Therapy is offered in two situations, say occupational therapy (Situation 1) and speech therapy (Situation 2). A behavioural programme is devised, aimed at reducing the frequency of the aggressive outbursts. A multiple baseline design involves the observation of the frequency of the target behaviour in both Situations 1 and 2. After establishing a baseline, the intervention is introduced first in one of the situations and then in the other. Evidence demonstrating the effectiveness of the behavioural intervention is shown in Figure 7.5.

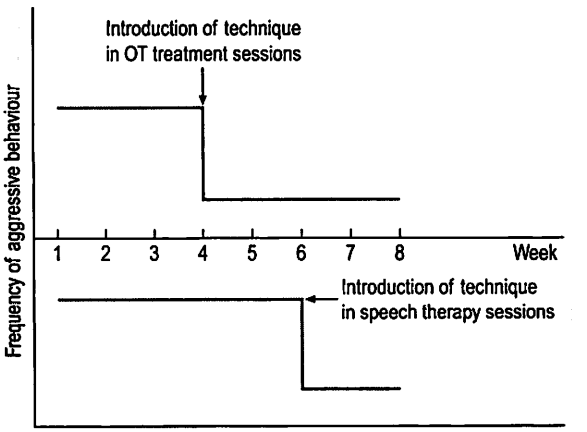


Figure 7.5 • Aggressive behaviour of a brain-damaged patient in two situations.

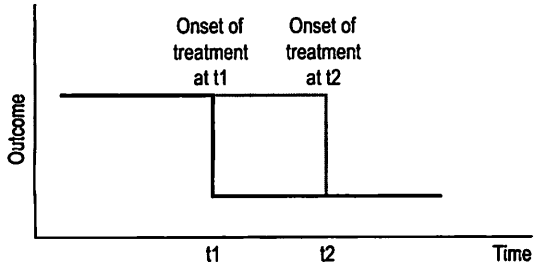


Figure 7.6 • General structure of multiple baseline designs.

The hypothetical data presented in Figure 7.5 indicates that the frequency of the target behaviour (aggression) declined in Situation 1 when the intervention was introduced, while it remained stable in Situation 2. The subsequent introduction of the intervention in Situation 2 resulted in a decrease of the behaviour.

Multiple baseline designs can also be introduced by generating baselines for two or more behaviours, or for two or more individuals. Just as in the example involving the different situations, the interventions would be introduced first for one of the behaviours or individuals and, subsequently, for the others. Figure 7.6 illustrates a general example for multiple baseline designs.

Clearly, by introducing the intervention at different times, we are controlling for the effects of extraneous variables that might have influenced our dependent variable. Both ABAB and multiple baseline designs are appropriate for demonstrating the benefits of therapeutic procedures in individual patients.

The interpretation of the results for $n = 1$ designs

The principles for conducting and interpreting health research are also relevant for evaluating the effectiveness of everyday health practices. The need for control and accurate measurement arises when we intend to demonstrate that particular interventions or preventive interventions are causally related to beneficial outcomes.

As we have seen previously we can exercise control by systematically introducing and withdrawing interventions and measuring concomitant changes in the signs and symptoms of a disorder.

Hypothetical example

Let us examine a hypothetical example for illustrating four different methodological principles relevant to the interpretation of $n = 1$ studies. Imagine that you are caring for a young man called John who had suffered a severe fracture of the femur following a motorbike accident. He is gradually recovering with intensive rehabilitation but he is still in severe pain when narcotic analgesics are not provided.

Previous research has shown that transcutaneous electrical nerve stimulation (TENS) was a safe and effective modality for managing acute pain. However the intervention is not equally effective for all patients so that there are no absolute guarantees that TENS will control pain in this particular patient. Also, even biologically inert pain control techniques can reduce pain through placebo effects. In this case you decide to investigate two clinically relevant research questions:

1. TENS is effective in reducing John's pain.
2. The pain reduction is causally related to the action of TENS.

Having obtained your patient's informed consent you decide to conduct an ABAB, $n = 1$ design for obtaining evidence to answer the above research questions. The variables here can be defined as follows:

1. *IV.* The independent variable is the pain control strategy using the TENS machine. The electrical stimulation produced by the TENS machine can be regulated such that you are in a position to vary the output of the machine. Two settings are selected:
   (a) A represents a TENS output which is known to be physiologically *ineffective* for activating the 'pain-gating' mechanisms (see Section 1). A represents the 'placebo control' level of the IV.
   (b) B represents a TENS output which is known to be physiologically *effective* in activating 'pain-gating' mechanisms. B represents the 'intervention' level of the IV.

2. *DV*. The outcome or dependent variable is an explicit measure of aspects of the signs and symptoms of the patent's condition. In this case we can select 'pain intensity' as a relevant dependent variable, measured on a 10-point scale (on this scale 0 represents 'no pain at all' and 10 'excruciating pain').
3. *EV*. Extraneous variables are variables other than the IV which can influence the DV and may 'confound' the demonstration of causal effects (Ch. 4). In the present $n = 1$ study we are controlling for extraneous variables by:
   (a) using a placebo control condition
   (b) introducing and withdrawing the active intervention over time (ABAB).

The therapeutic effectiveness of the TENS for pain control will be represented by the changes in the pain intensity or the differences between the A and B phases of the intervention. Table 7.1 provides the procedure for conducting the data collection.

The 'A' phase gives us the baseline against which we can compare the effectiveness of the active TENS intervention for reducing pain intensity. The outcome indicates a large and consistent difference between the A and B levels of the independent variable, and the large difference between A and B indicates a clinically useful reduction in the pain intensity experienced by John, your patient. You are now justified in continuing the intervention until he is sufficiently recovered to be painfree.

We hope you will agree that this was an inspiring little story, the problem with it being that things just do not normally happen this way in the real world. The data we hypothesized above clearly illustrate an ideal, easily interpreted outcome appropriate for educational purposes. However, there may be some factors in the real world that impact upon these idealized findings. These include:

1. *Variability*. The above results are idealized, as opposed to 'variable'; that is to say, appearing to be all over the place. We will look at the concept of variability in Chapters 14 and 15. At this stage it suffices to say that the more variable the findings, the more difficult it is to identify the trends and differences in results of $n = 1$ studies.
2. *Effect size*. The effect size is indicated by the difference between the baseline (A) and the intervention (B) phases. If the effect size is large in relation to the variability of the data, changes in pain intensity will be easy to see. However, if the effect size is small in relation to the variability of the data, then changes in pain intensity will be hard to see. Thus, the ability to detect the effects of interventions is affected by natural variability in the measured variable.

Table 7.1 Data collection procedure for hypothetical example

Time period	Independent variable level	Dependent variable
Week 1	A phase (placebo)	Daily assessment of pain intensity (at 0800 hours)
Week 2	B phase (treatment)	
Week 3	A phase (placebo)	
Week 4	B phase (treatment)	

The validity of $n = 1$ designs

We examined three (AB, ABAB, multiple baseline) designs in the previous subsections. However, more complicated, 'mixed' designs are available for $n = 1$ investigations. The mixed designs include elements of both reversal and multiple baseline strategies. We will not discuss these in detail in this text; interested readers are referred to Barlow & Hersen (1984).

A basic requirement for the valid interpretation of all $n = 1$ designs is the production of a stable baseline. Unless this requirement is met, the interpretation of the results is extremely difficult. In some clinical situations, the production of a stable baseline might be unethical as it could involve withholding treatment. The intervention phase must also be long enough for the effectiveness of the treatment to emerge. Some interventions, such as those involving physical rehabilitation or psychotherapy, might need to be administered for months before their effectiveness, or lack of it, becomes apparent. Clearly, we cannot assume that the baseline and intervention phases are of equal duration.

It is essential that the observations should be valid and reliable. Some observations are straightforward, such as those based on taking temperatures. However, given a more complex variable such as 'aggression' we need to establish with clarity that different observers agree on the type of behaviours we are going to observe; behaviours that might seem 'aggressive' to one observer might not be classified as such by another.

We have already discussed how $n = 1$ designs attempt to control for the influence of extraneous variables. Although the $n = 1$ designs can be conceptually adequate to demonstrate causality, the patients, being in their natural setting, can be influenced by all sorts of uncontrolled events. After all, it is not possible to insulate individuals from their environment. Therefore, sources of invalidity must be evaluated with respect to each $n = 1$ investigation. It must also be remembered that no matter how sophisticated the $n = 1$ design, the observed outcomes for any given case may not generalize to other cases.

Summary

In this chapter, we examined three (AB, ABAB and multiple baseline) of the designs available for studying single individuals in their natural settings. It was argued that ABAB and multiple baseline studies provide a valid means for evaluating the causal effects of variables on therapeutic outcomes. These $n = 1$ designs are particularly useful for establishing the usefulness of interventions for individual patients. Although some limitations and ethical constraints might emerge in conducting $n = 1$ studies, they provide a useful tool for practising health professionals interested in evaluating the effectiveness of their interventions.

Although the statistical analysis of $n = 1$ studies is beyond the scope of this introductory text, it should be noted that graphing our observations, as discussed in this chapter, provides evidence for possible causal relationships. A precondition for interpreting the results of $n = 1$ studies is having an adequate number of stable observations across the various conditions.

The $n = 1$ designs are quite similar to quasi-experimental designs, in that the investigator has control over the type and timing of the interventions.

Self-assessment

Explain the meaning of the following terms:

AB design
ABAB design
baseline
designs reversal
multiple baseline

True or false

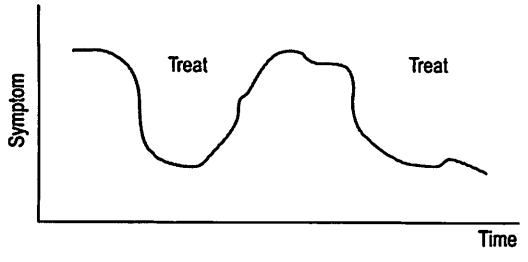
1. $n = 1$ designs are most appropriate for conducting epidemiological investigations.
2. The graphing of $n = 1$ findings involves graphing the magnitude of the dependent variable along the y-axis and time or sessions along the x-axis.
3. With an AB design, the time period A represents the time period during which the intervention is introduced.
4. A baseline is a series of observations recorded during a period when the intervention is administered.
5. With an AB design an intervention is introduced, withdrawn and then re-introduced.
6. A limitation in using ABAB designs in clinical investigations is that it might be possible, but unethical, to withdraw a treatment.
7. Using multiple baselines across two situations involves establishing baselines, then administering treatments simultaneously in both situations.
8. An advantage of using multiple baseline designs over AB designs is the increased control over threats to internal validity.
9. AB designs involve a period of initial treatment followed by the withdrawal of the treatment.
10. For $n = 1$ designs, the 'B' period is when the treatment is administered.
11. $n = 1$ designs cannot, in principle, indicate causal effects.
12. The major advantage of $n = 1$ designs is that no ethical issues need to be taken into account in planning such research.

13. An ABAB design involves introducing the treatment twice.
14. An ABAB design provides a more powerful control over possible extraneous variables than an AB design.
15. ABAB designs are particularly useful when changes in the symptoms following treatment are irreversible.
16. The production of a stable baseline might be unethical in some clinical situations.

Multiple choice

1. Quasi-experimental and $n = 1$ designs are similar in that:
   a they can both involve time-dependent sets of observations
   b both can involve control over the time at which a treatment is introduced
   c both *a* and *b*
   d neither *a* nor *b*.
2. An ABAB design:
   a involves the use of two individuals, A and B
   b has two baseline periods
   c depends on manipulating A and measuring B
   d involves correlating AA with BBC.

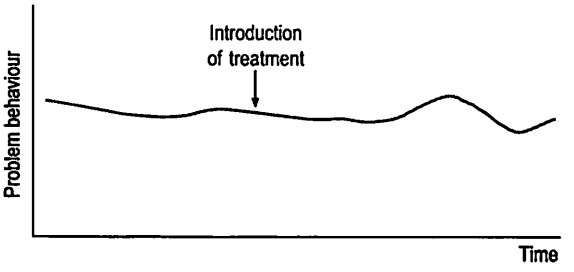
Consider the following outcome for an $n = 1$ type investigation.



Questions 3 and 4 refer to the graph above.

3. This is a(n):
   a AB design
   b ABA design
   c ABAB design
   d 'mixed' design.
4. The above graph indicates that:
   a the treatment appears effective in reducing the symptom
   b the treatment appears to exacerbate the symptom
   c the treatment appears to have no effect on the symptom
   d no conclusion can be reached on the basis of the obtained data.

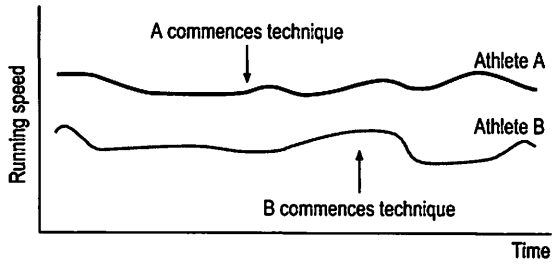
A child with a serious behavioural problem at home is institutionalized. After a 1-week period of observation, a behavioural treatment is introduced. The graph below represents the fictitious results.



Questions 5 and 6 refer to the above data.

5. This is a(n):
   a AB design
   b ABA design
   c ABAB design
   d multiple baseline design.
6. The above graph indicates that:
   a the treatment appears to have improved the child's problem behaviour
   b the treatment appears to have exacerbated the child's problem behaviour
   c the treatment appears to have had no effect on the child's behaviour
   d a factorial design should have been used.
7. Treatment is undertaken to attempt to reduce anxiety in an individual. An ABAB design is used to evaluate treatment effectiveness. Because of different shifts at a hospital, six different observers are used to record the anxiety-related behaviours. Problem(s) with this investigation is (are) that:
   a the observers might not record exactly the same behaviours as reflecting 'anxiety'
   b the presence of the different observers might influence the behaviours in different ways
   c both *a* and *b*
   d neither *a* nor *b*.

A new technique to improve running speed in athletes is evaluated. The graph below represents fictional results for two athletes.



Questions 8 and 9 refer to the above data.

8. This design is:
   a AB
   b ABAB
   c multiple baseline
   d factorial.

9. The above graph indicates that:
   a the new technique appears to improve running speed
   b the new technique appears to reduce running speed
   c the new technique appears to have no effect on running speed
   d further investigations on the effects of the technique are necessary.

10. The use of an ABAB design is most useful for reducing threats due to:
    a placebo effects
    b experimenter expectancy
    c maturation
    d assignment.

Chapter Eight

8

Qualitative research

CHAPTER CONTENTS

Introduction

The research strategies discussed in previous chapters can be called 'quantitative' in that the data obtained consist of measurements which can be statistically analysed. Quantitative research provides a particular perspective on health and illness. However, there are alternatives. Qualitative or interpretive research involves the investigation of individuals and groups in their social settings. The investigator seeks to understand the thoughts, feelings and experiences of the research participants as people coping with their conditions and treatments in a given social setting. The use of evidence from qualitative studies has traditionally been a fundamental source of knowledge in the clinical and social sciences.

The aims of this chapter are to:

1. Outline different conceptual approaches to qualitative research.
2. Compare and contrast specific dimensions of qualitative and quantitative approaches to research.
3. Emphasize the importance of using evidence from both qualitative and quantitative research.
4. Discuss basic strategies used for ensuring the validity of qualitative research.
5. Examine the scope and limitations of qualitative research in the health sciences.

What is qualitative research?

Qualitative research is disciplined enquiry examining the personal meanings of individuals' experiences and actions in the context of their social environments. 'Qualitative' refers to the nature of the data or evidence collected. Qualitative data consist of detailed descriptions based on language or pictures recorded by the investigator. By 'disciplined' we mean that the enquiry is guided by explicit methodological principles for defining problems, collecting and analysing the evidence, and formulating and evaluating theories.

'Personal meaning' refers to the way in which individuals subjectively perceive and explain their experiences, actions and social environments. Qualitative research provides systematic evidence for gaining insights into other persons' views of the world, 'putting ourselves into someone else's shoes'.

There are a variety of approaches to qualitative research and these take different positions concerning how data should be collected and analysed. There are also several diverse schools of thought that have contributed to the historical development of qualitative field research (see, for example, Denzin & Lincoln 1994, Liamputtong Rice & Ezzy 1999).

1. *Phenomenology*. Phenomenology, which is both a system of philosophy and an approach to psychology, emphasizes the direct study of personal experience and the understanding of the nature of human consciousness. Research in this area involves 'bracketing' or putting aside the usual preconceptions and prejudices that influence everyday perception so that we can uncover the pure constituents of conscious experience. Within this framework, conscious experience is seen as the basis for personal meaning as we reflect on our experiences in the context of our goals and purposes. An important concept adopted from phenomenology is the notion of 'multiple realities', that is, different people may consciously experience the world in quite diverse ways. This suggests that in order to understand the meanings of a person's actions, we must become adept, through empathy, at seeing things from their point of view.
2. *Symbolic interactionism*. Symbolic interactionists emphasize that a social situation has meaning only in the way people define and interpret what is happening. That is, people do not react to 'objective' aspects of their environments, but rather their actions are guided by their personal interpretations of the situation. It follows that different people, on the basis of their past experiences and their particular social positions, may come to interpret a specific situation in quite divergent ways, and act in conflicting fashions. For example, a male obstetrician might view childbirth in quite a different way from a female midwife, and in turn their views might be quite discordant with that of a woman giving birth. In social and health care settings, it is useful to explore different perceptions of events, as it is clear from the work of symbolic interactionists that 'shared perspectives' among people cannot be taken for granted.
3. *Ethnomethodology*. Ethnomethodologists study the processes associated with the way in which people perceive, describe and explain the world. Ethnomethodologists argue that the meanings of specific actions and events are not necessarily obvious, but are in fact rather ambiguous and problematic. People select and apply specific rules and principles in order to define and give meaning to situations in which they find themselves and in order to justify their actions in a given situation. Ethnomethodologists assert that we take an enormous amount of cultural context, such as norms and rules, for granted in our everyday communications and social interactions, and we tend to 'bracket' this as obvious or common sense. It must be remembered, however, that when the cultural backgrounds of individuals diverge, the understanding of personal meaning becomes less obvious or commonsense.

Although taking somewhat different views of personal meanings, the above three approaches have common themes and have all contributed to the development of qualitative field research. Table 8.1 shows key aspects of qualitative field research, in contrast to quantitative approaches.

Data collection and interpretation in qualitative field research

The fundamental aim of planning and designing qualitative field research is to position the investigator close to the participants, so as to gain access

Table 8.1 Contrast between quantitative and qualitative methods

	Quantitative	Qualitative
Perception of subject matter	Reductionistic: identification and operational definition of specific variables	Holistic: persons in the context of their social environments
Positioning of researcher	Objective: detached observation and precise measurement of variables	Subjective: close personal interaction with subjects
Database	Quantitative: interrelationships among specific variables	Qualitative: descriptions of actions and related personal meanings in context
Theories	Normative: general propositions explaining causal relationships among variables	Interpretive: providing insights into the nature and social contexts of personal meanings
Theory testing	Controlled: empirically supporting or falsifying hypotheses deduced from theories	Consensual: matching researcher's interpretations with those of subjects and other observers
Applications	Prediction and control of health-related factors in applied settings	Interacting with persons in a consensual, value-consonant fashion in health care settings

to and describe personal experiences, and to interpret their meanings in specific social settings. The following subsection develops in more detail the corresponding points presented in Table 8.1.

Perception of subject matter

Qualitative research is preferably carried out in a natural setting and there is no attempt made by the investigator to control for extraneous influences. Furthermore, there are no operational definitions provided for the study variables, but rather the health-related experiences being studied are perceived and described as a whole, in their social contexts.

Strong preconceptions or fixed hypotheses are not advantageous for qualitative field research. This is a different situation from that in quantitative research, where there are precisely defined hypotheses or aims for guiding the research. Qualitative researchers do have general aims and theoretical notions pertaining to the phenomena being studied but these are tentative and are open to modification as the data collection proceeds.

Qualitative field research focuses on the in-depth understanding of specific individuals and groups, rather than studying the general characteristics of a large population of individuals across specific variables. It should be kept in mind, however, that some quantitative designs may address single cases rather than large study populations. For example, we have reviewed $n = 1$ designs. The difference is that $n = 1$ designs address specific variables representing aspects of the individual's behaviour or clinical symptoms, rather than attempting to describe and understand individuals holistically in the context of their natural social settings. Such an approach is called 'idiographic' (describing a specific individual) as opposed to the 'nomothetic' (describing general phenomena) view of research in quantitative research.

Positioning of researcher

Accurate and replicable measurements are valued in quantitative research. The fundamental positioning of the researcher is 'objective', that is, aiming to perceive and record events without any personal

bias or distortion. Within this quantitative framework we assume an objective reality that can be disclosed through accurate reproducible measurement or observation. The situation in qualitative field research is far more complex, as the researcher is more a part of the phenomenon being investigated than the detached observer in quantitative research. To understand personal meanings and subjective experiences one has to become involved with the lives of the subjects being studied. That is, some degree of empathy must develop between the researcher and the subject. By empathy we mean the ability to 'put ourselves in the other person's shoes' or to see things from their perspective(s). Reality is said to be constructed by the individual.

A particular reason for the advancement of quantitative research has been the development of valid and reliable measurement instruments. However, when standardized tests and measures are used to study a person, they become 'enframed' within the limitations of the instrument, and their possible unique self-expression may remain outside the scope of the enquiry. The qualitative researcher may find instruments intrusive, restricting the possibilities for understanding the ideas and emotions of the respondent.

There are advantages to a 'human measuring instrument' which is used in qualitative research. After all, we are more adaptable and multi-purpose than even very sophisticated machinery and we may be able to observe subtle behavioural changes and verbal and non-verbal cues in our participants. In addition, as the investigation progresses, the human 'instrument' becomes more aware of what is happening and as we tune in with each other's points of view the data collection becomes more accurate.

Database

The data obtained in quantitative research consist of sets of measurements of objective descriptions of physical and behavioural events. These are summarized and analysed in accordance with statistical principles outlined at an introductory level in Sections 4 and 5 of this book. The data in qualitative research are descriptive, a 'thick' or thorough description of what people said, their actions and activities, non-verbal behaviours and interactions with other people: 'The reality of the place should be conveyed through representation of its mundane aspects in a straightforward manner' (Lofland 1971, p. 4). An important aspect of field research is keeping thorough, up-to-date field notes. These should be recorded as closely as possible to the time of occurrence of the phenomena under study. The field notes should contain direct quotations from the participants and the settings in which the statements and actions were recorded. Where possible (where it is appropriate and not overly intrusive), the researcher may use audio and video recordings. This helps to record interviews, and improves accuracy in conveying what was said and done in a given setting, since it is possible to review the obtained information.

Although 'objectivity' does not mean remaining detached from the situation, it is essential in qualitative research that the reports of events should be truthful. The investigators should not allow ideological biases to distort or censor their observations, or deliberately lie to place their subjects in a good or bad light. This is a particularly important point as, given the close personal interaction with the subjects, one may be predisposed to report favourably.

By 'database', we mean the overall evidence that forms the basis for theory formation and specific applications for health care. In quantitative research, the database will consist of the statistically treated data which will enable us to see how specific variables are interrelated. In terms of qualitative field research, the database is essentially a narrative (or a story, if you like) that reports what has happened to people, what they did or said in specific situations. This narrative should be adequately detailed so as to illuminate for the reader the personal meanings that the health-related events had for the informants.

Theories

Theories represent our current state of knowledge about the state of the world. Theories are abstract, coherent explanatory systems which integrate a broad range of research findings. Theories may be

constituted of premises stated in everyday language, with particular attention paid to the appropriate use of concepts and the logical development of the premises.

Theories based on quantitative evidence integrate patterns of findings concerning the interrelationships among variables. Such theories often contain 'models', which may be mathematical and/or systems representations of the patterns of findings. Models of anatomical and physiological processes, such as those of the circulatory or nervous systems, are good examples of successful quantitative models. Conversely, theories integrating evidence from qualitative research do not address facts about how objects are constituted and interact, but rather are the overall interpretations of personal meanings emerging in specific social settings.

Some commentators (Guba & Lincoln 1983) argued that data collection and theory formation should be intrinsically integrated rather than being different stages of the research process. In addition, it is suggested that personal meanings should be seen as unique and idiosyncratic, and thus no attempt should be made to integrate systematically such diverse personal positions. Theory, from an idiographic position, is seen essentially as the accurate presentation of the situation from a particular person's perspective.

Other qualitative researchers approach theory formation by attempting to identify common 'themes' or categories of meanings emerging from the data. The important point here is that the theoretical categories are developed from evidence expressing personal meanings, rather than 'facts' derived from the statistical treatment of objective measurements concerning variables. In this way, theory is said to be 'grounded' in the narratives of particular individuals.

Some researchers stress the broad, culture-interpreting aspects of qualitative field research. The formation of critical theory explains how personal meanings and actions emerge and are influenced by the person's social and cultural milieu. Critical theories identify the extent to which individuals' self-perception and freedom for action may become distorted and limited by the operation of power and coercion within a culture (e.g. Grundy 1987, pp 15–19, 106–114). As a general illustration, critical theories of the lives and experiences of Western women in the 1960s were crucial to the development of feminist movements.

Theory testing

Theories based on quantitative evidence lead to clear-cut, empirically testable predictions or hypotheses logically deduced from the theories. Theories are supported or falsified by a body of evidence collected under controlled conditions. Testing qualitative theories is somewhat different, as no causal mechanisms are included in the theoretical framework. The simplest verification of qualitative interpretations is to go to the subjects themselves, in order to establish if the researcher's interpretations make sense to them. The extent to which a consensus develops between researchers and their subjects is one of the important indications of the truth of qualitative theories.

Applications in health care delivery

The applications of quantitative evidence and theories are essentially technical, providing mechanisms in terms of which we can predict and control specific health-related variables. That is, we apply quantitative approaches for discovering the causes and progress of diseases and disabilities, for developing and validating assessment procedures and for evaluating the effectiveness of interventions.

In contrast, qualitative field research provides evidence and theories that enable us to understand our clients better as human beings. This research discloses how illnesses, disability and health care delivery affect people's lives interpreted from their points of view. In the following subsection, we will examine some of the applications of qualitative field research for improving health care delivery.

Qualitative field research

When there are significant differences in the cultural backgrounds and experiences of persons, the understanding of personal meanings becomes problematic. For example, an anthropologist might need

to spend decades immersed in, and systematically studying, a different culture to be in a position to interpret accurately the actions and traditions of the participants.

There are numerous areas of health research where involving the interpretation of personal meanings is essential to ensure effective practices. The following three examples illustrate areas where qualitative field research can make strong contributions for clarifying personal meanings.

1. Understanding cultural differences between health workers and clients. In countries such as the US, Australia, Canada or the UK we live in multicultural societies. There is persuasive evidence that the way people experience their bodies, or events such as childbirth, pain or illness depends to a large extent on their cultural backgrounds. When health practitioners misconceive their clients' view concerning their illness or injury, the outcome may be erroneous diagnoses and useless interventions. A particularly important area of qualitative field research is to clarify personal meanings of clients and therapists with regard to health care problems, in an attempt to improve communications and enhance treatment outcomes.
2. Evaluating the effects of health care environments. Health care institutions, such as general and mental hospitals, can be seen as 'subcultures' having strong influences on the lives of both staff and clients. Persons with chronic illnesses and disabilities requiring long-term care might come to view themselves and their life situations from an 'institutionalized' perspective. The development of critical theories in these areas is particularly relevant for understanding the influences of health care environments. Research findings in this area have been applied to devise strategies to empower people such as those with intellectual disabilities to live and participate in the community.
3. Relating to people with neurological or psychiatric problems. People diagnosed as suffering from problems such as schizophrenia, intellectual disability or brain disorders may, to some extent, experience themselves and the world in ways different from 'ordinary' people. How such persons experience aspects of their world is by no means obvious, as these clients may demonstrate severe information-processing impairments such as delusions, hallucinations or memory problems, which may make it extremely difficult to establish empathic relationships. However, in order to ensure that persons with such severe impairments or disabilities are treated appropriately and with understanding, health professionals must learn to see things from their perspectives. Qualitative field research has provided evidence which has helped to clarify the personal perspectives of people with severe disabilities.

The above are some obvious examples where qualitative research is appropriate for clarifying personal meanings, and enhancing understanding and communication in health care settings. However, personal meanings are relevant to all health care situations, not only in the obvious areas discussed above. The following exemplify questions which are appropriately approached through field research strategies:

- What is it like to have a speech disorder? In what ways does it disrupt the person's life, from their points of view?
- How do caregivers interact with terminally ill patients? How do health professionals experience the death of a patient?
- How do health professionals break the news of unfavourable diagnoses, such as heart disease, to their patients? How are such situations seen from the perspectives of the health professional or the patients?

The integration of quantitative and qualitative methodologies

When used jointly, quantitative research tools can be particularly powerful. One of the authors has conducted a study of how people evaluate primary health services (Thomas et al 1993). The first step in this process was to conduct focus group interviews with 20 groups of 8 participants specifically selected from a wide range of ethnic backgrounds, ages and sexes. The groups were conducted by a facilitator who presented questions concerned with knowledge and opinions of, and satisfaction with, health services. The discussions were recorded and transcribed.

One set of analyses of the transcripts involved consideration of everything that had been said about the health services with regard to satisfaction

or dissatisfaction. This resulted in a range of separate categories or themes. These themes, therefore, were directly derived from the participants' own words and interpretations of their experiences.

The themes were then framed in the form of questions that sought information from people about their satisfaction and dissatisfaction with health services. The questions were then incorporated into a questionnaire (see Ch. 9). When the questionnaire was piloted with a sample of 500 people who attended several doctors' surgeries over a period of three weeks, it was found that none of the participants nominated new factors that affected their satisfaction and dissatisfaction. Thus, the procedure used in developing the questionnaire had very effectively captured how people decided whether they were satisfied or dissatisfied with their health services. This study is an example of where qualitative and quantitative research methodologies can combine powerfully. There are many productive ways for combining quantitative and qualitative approaches to health research. Interested readers are advised to consult Tashakkori & Teddlie (1998).

The validity of qualitative field research

It should be noted that the unstructured and descriptive nature of the data collection process in field research often sits uneasily with those favouring 'quantitative' research strategies. The major problem with the unstructured data collection techniques is that observer bias may cloud or distort the data being collected. As previously discussed, there are well-known observer effects such as the Rosenthal effect and the Hawthorne effect. Structured data collection methods are most likely to control for these effects, although there are no guarantees that they will be eliminated.

Furthermore, the sampling processes involved in qualitative field research are complex. Most social phenomena are profoundly affected by their participants. 'Real' situations may not reflect these biases. An important issue in understanding qualitative research is the specific culture dependence of the findings; what is true in one social setting may not be true in another. Also cultures change with time. For example, the experiences of psychiatric patients who lived in large closed mental hospitals up to the 1970s might not be generalizable to mentally ill people who nowadays live in the community.

Therefore, as in other types of research, qualitative field studies also have to confront problems of external and internal validity. Guba & Lincoln (1983) recommended a variety of strategies to ensure the validity and reliability of field studies. These strategies included:

- asking subjects if the observations about them are credible (believable)
- prolonged engagement by the observers to minimize distortions caused by their presence
- triangulations, which involved pitting against each other different data and theoretical interpretations to provide cross-checks of observations and interpretations.

There are many other ways for ensuring the validity or rigour of qualitative research. These more advanced issues can be explored in books dedicated to qualitative research in health care (e.g. Liamputtong Rice & Ezzy 1999).

Therefore, despite controversies in the area, qualitative researchers pay considerable attention to methodological issues to ensure the adequacy (that is, validity and reliability) of their investigations. The situation is essentially no different from quantitative research, although qualitative researchers take somewhat different steps to ensure the accuracy and generalizability of their findings.

Summary

Qualitative research strategies include data collection which is aimed at understanding persons in their social environments. Rather than generating numerical data supporting or refuting clear-cut hypotheses, field research aims to produce accurate descriptions based on face-to-face knowledge of individuals and social groups in their natural settings. The role of the observer in this context is crucial and usually involves physical and social closeness between the subject and the observer. Data collection involves objective and accurate

reporting of the activities and appearances of persons in their natural environments. As with other strategies of research, investigators must pay considerable attention to the external and internal validity of field research. We briefly looked at some ways in which field researchers can cross-check their descriptions in an attempt to ensure the validity of their reports and interpretations.

Different research designs may be used to generate evidence of the same processes, although from different perspectives. For instance, any complex clinical phenomenon, such as schizophrenia, may be studied using any of the research strategies outlined in Chapters 4–8. To understand the scope of the problems and the effectiveness of the appropriate treatments, it is desirable to use a variety of research strategies. Conversely, a comprehensive theory of a clinical problem should generate any number of hypotheses within the realm of the research strategies discussed in this book. We will look at the analysis of qualitative data in Chapter 21.

Self-assessment

Explain the meaning of the following terms:

critical theory
empathy
ethnomethodology
personal meaning
phenomenology
qualitative field research
quantitative research

True or false

1. Qualitative researchers focus on individuals behaving in their natural environments.
2. According to Lofland (1971), the field researcher should maintain a 'distance' from subjects in a physical and social sense.
3. A problem with unstructured data collection techniques is the possible distortion of the evidence through observer bias.
4. The basic aim of qualitative research is to test clearly defined hypotheses.
5. It is essential in qualitative research to place disadvantaged or oppressed people in a 'good light', to further their needs or causes.
6. Qualitative researchers need not concern themselves with issues of validity.
7. Qualitative research generally involves the use of precision instruments to measure specific subject variables.
8. Qualitative research generally produces intimate, face-to-face knowledge of other individuals.
9. Phenomenologists are concerned with the understanding of the nature of human conscious experience.
10. Quantitative research produces data well suited to the formulation of causal models.
11. Quantitative research is most appropriate for interpreting personal meanings in social settings.
12. Quantitative research is best suited to discovering how biological systems work.

Multiple choice

1. Qualitative research involves:
   a the testing of clear-cut hypotheses by employing sophisticated measuring instruments.
   b empathy with subjects' points of view
   c structured data collection
   d carrying out research in the open air.
2. Which of the following is not an example of qualitative research?
   a A researcher studying nursing conditions in major hospitals spends a week working as a nurse aide at Prince Henry's Hospital.
   b An anthropologist goes to live with a New Guinea tribe to find out about their religious practices.
   c A psychologist studying therapeutic processes attends group therapy as a client.
   d A speech pathologist compares two rival methods of treatment for stuttering.
   e A physiotherapy student spends a day in a wheelchair and uses this experience to write a report on some of the problems of the physically handicapped.
3. Which of the following disciplines would most likely employ qualitative designs?
   a nuclear physics
   b anatomy

c genetics
d sociology.

4. The medical model, as discussed in Chapter 1, is best supported by:
   a qualitative field research
   b quantitative research
   c philosophical speculation
   d all of the above.
5. Which of the following is not a characteristic of quantitative research?
   a A holistic approach to persons.
   b Precise definition of variables being studied.
   c Prediction and control of phenomena.
   d Theories including causal models.
6. An important basis of qualitative field research is:
   a phenomenology
   b numerology
   c measurement theory
   d the medical model.
7. A psychiatrist is interested in research to identify the relationship between brain dopamine levels and the occurrence of specific well-defined abnormal behaviours. This research would be:
   a based on phenomenological principles
   b a project in an ethnomethodological framework
   c a quantitative project
   d best described as qualitative field research.

Section Three

Discussion, questions and answers

The health sciences clinical researcher has a rich variety of research designs from which he or she may choose, including the experimental survey, single case and qualitative field approaches. Each approach has its unique strengths.

The experimental and single case designs are particularly suited to studies of the effects of health interventions. These are studies that are concerned with the clinical impact of administering different types of interventions (or non-interventions) upon people. In the classical two-group experimental study, similar groups of people are administered with either an intervention or a non-intervention condition and their outcomes are compared. The use of non-intervention with people who are ill has moral and ethical implications. So, often the experimental method is used to study the relative effects of two or more different types of intervention. There has been an affinity between the professions where health interventions are the norm and a form of enquiry which also involves intervention (i.e. the experimental and single case types of research design). However, much high-quality health sciences research involves other types of research design.

Sample surveys are frequently employed to study the opinions of large groups of people concerning health services and their experiences of health and illness. Epidemiological research, which is aimed at studying the distribution of health and illness and associated risk factors in populations, generally involves large-scale sample surveys. These surveys often draw upon hospital and other health agency records.

Qualitative field research involves the disciplined examination of the personal meanings of individuals' experiences and actions in the context of their social environment. The emphasis in such research is upon depth of interpretation rather than extent of sampling. Therefore such studies typically employ much smaller samples of participants than, for example, sample surveys.

It is useful to consider examples of the application of the various techniques to the same type of research questions.

Let us consider the situation of people with back injuries associated with manual labour. The major problem associated with back injuries is disability arising from pain. In countries such as the US and Australia older workers who come from non-English-speaking backgrounds are over-represented among people with these injuries. This is probably because such people are over-represented in jobs that have a greater risk of back injury and the effects are cumulative over a long period.

Experimental example

Let us consider an example of where a clinical researcher is interested in comparing the

effectiveness of two alternate interventions for the treatment of back injury. There are two common approaches: a conservative approach such as physiotherapy, and surgery. So, if we took 20 workers with back injuries and, after random allocation, 10 of them were treated with a conservative intervention such as physiotherapy, and 10 were treated with surgery, we could compare their outcomes.

To achieve a true experimental design, the people need to be randomly allocated to surgery or physiotherapy. Incidentally, to implement this procedure in real life would require a lot of talking to the relevant ethics committees. All the participants would have to volunteer.

As far as the measurement of the outcome is concerned, most experimental studies would employ a quantitative measure of outcome. For example, the workers could fill in a pain questionnaire, perhaps on several occasions after the interventions, in order to compare the outcomes for the two groups. The use of a written questionnaire, especially if it is provided in English only to people from a non-English-speaking background, makes assumptions about the literacy of the participants which may not be valid.

Natural comparison example

If the clinical staff had chosen the interventions for the patients on a non-random basis, the study as described above would be a natural comparison study. Simply comparing groups does not mean we have an experimental design. Otherwise, the study could be structured in an identical fashion to the experimental example described above.

Single case example

A single case design involves the administration and withdrawal of interventions in a systematic fashion in order to observe their impact upon the phenomenon under study. The person in the single case trial is usually challenged with varied interventions to compare the effects of each. This type of design closely approximates the natural clinical history of interventions with many patients, particularly those with chronic illnesses. However, the interventions in a single case study are structured much more rigorously.

In the present context, while it would be possible to administer and withdraw physiotherapy and to have baseline phases of no treatment, the surgical intervention cannot be withdrawn. Once the surgery is done, it permanently and drastically changes the person's body. Therefore a single case example could involve the alteration of baseline (no treatment) and physiotherapy and then, as the last link in the chain of event, the surgery. This is not a methodologically strong design, as there could be carry-over effects from the previous interventions that interact with the surgery, and the order of the administration of the surgery could not be readily altered.

Survey example

An alternative way of conducting the study of workers with back injuries would be to select a large group of such people and survey them for the different types of interventions they have had for their injury. The outcomes for different groups could then be compared because, as in a natural comparison study, we can compare men and women respondents. We could even give them the same pain questionnaire as in the types of study previously discussed.

Surveys do not, of course, necessarily involve the use of questionnaires. One could bypass the patients and survey their medical records (with necessary approvals) using a coding schedule. Alternatively, the information could be collected in the form of a structured survey interview.

Qualitative field research example

An additional way of studying this issue would be to conduct in-depth interviews with a small number of injured workers to study their interpretations of their situation. Workers who had surgery, as well as those who had physiotherapy, could be interviewed. The interviews would normally be recorded verbatim, transcribed and then exhaustively analysed for the theoretical constructs needed to describe the experiences of the participants. The use of checklists, survey

inventories, etc. is normally avoided, although it is usual to have a list of issues to be introduced by the interviewer. The respondents typically describe their experiences and perceptions in their own natural discourse.

In studies which use questionnaires, there is limited interaction between the researcher and the respondent. As suggested by their name, the respondents respond to the questions framed solely by the researcher. Such an approach can be useful in eliciting factual information in an economical manner. However, it places the respondent in a relatively passive role. If the researcher and the respondent do not share the same constructs, ideas, feelings and motives, the 'wrong' questions may be asked. Schatzman & Strauss (1973, p. 57) note that the researcher 'harbours, wittingly or not, many expectations, conjectures and hypotheses which provide him with thought and directives on what to look for and what to ask about.' The respondent in most questionnaire studies has very little opportunity to contribute to the research agenda. Generally, a token 'any other comments' section is the extent of the invitation for the respondents to contribute to this agenda.

Questions

1. What would be the 'best' design to study the impact of physiotherapy and surgery upon the pain of workers with back injuries?
2. If the participant has been measured frequently over an extended time period, what type of data analysis is required?
3. What type of research design does an epidemiologist typically employ?
4. What special arrangements should be made for participants in research studies with low literacy in English?
5. Is it possible to use an experimental design with in-depth interviewing techniques?

Answers

1. No doubt this question will generate much lively debate. The authors' view is that there is no single best method to conduct such investigations. Each of the methods listed has legitimate insights to offer, each providing a perspective different from that of the others, provided the study is conducted well.
2. This type of design requires an analysis method that can handle repeated measures data, i.e. this is a repeated measures design.
3. Although they may use a variety of different designs, the epidemiologist is likely to use large-scale surveys, often of clinical records. The goal of epidemiologists is to determine the patterns of distribution of illnesses and their relationship with risk factors. Although the modelling may be complex, the research design is simple. Do a large survey.
4. Clinical researchers who are often fluent in English frequently do not make appropriate arrangements for participants with levels of literacy in English lower than their own. Such arrangements could include the provision of interpreters, translated instructions and questionnaires, and interviewers to read the questions to the participants.
5. Yes, it is possible to do so; however, it is very unusual. Experimenters typically employ quantitative outcome measures and researchers who use in-depth interviews tend to be disinclined to use experimental research designs. Perhaps both could learn from each other?

Section Four

Data collection

SECTION CONTENTS

There are numerous methods available for data collection. The appropriate methods are chosen depending on the aims, design and resources of the research project.

Questionnaires are commonly used with survey designs. In Chapter 9, we examine a number of ways in which we can draw up and validate questionnaires. There are different types of questionnaires, ranging from highly structured, standardized scales to unstructured open-ended formats.

An interview (see Ch. 10) is, in a sense, a conversation between the interviewer and the person being interviewed. As they require the presence of an interviewer, this increases the cost and effort needed to obtain data. The presence of an interviewer may also influence the respondent's answers. Qualitative research studies often employ in-depth interviewing techniques which are preferably carried out in the 'natural' settings (homes, hospital, etc.) in which the respondents are living or receiving treatment.

Observational methods are also commonly used strategies for data collection. They may range from highly structured observational protocols, indicating precisely which behaviours or clinical signs should be recorded, to unstructured records of the experiences of participant observers, as used in qualitative research.

Depending on how the data are recorded and analysed, interviews and observations may be used for both quantitative and qualitative research. However, the use of instrumentation to produce numerical data is most appropriate for quantitative research. A variety of standardized measurement instruments are now available for measuring biological and psychological functions (see Ch. 10).

Whatever data collection strategy is being used, we must ensure that it is reliable (replicable) and valid (accurate). Otherwise, as discussed in Chapter 12, the measurement error due to unreliable and invalid data collection strategies may prevent the researchers from achieving their goals.