Quantitative Methods for Health Research

A Practical Interactive Guide to Epidemiology and Statistics

Nigel Bruce, Daniel Pope and Debbi Stanistreet

Division of Public Health, University of Liverpool, UK

John Wiley & Sons, Ltd

Quantitative Methods for Health Research

Quantitative Methods for Health Research

A Practical Interactive Guide to Epidemiology and Statistics

Nigel Bruce, Daniel Pope and Debbi Stanistreet

Division of Public Health, University of Liverpool, UK

John Wiley & Sons, Ltd

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 6045 Freemont Blvd, Mississauga, Ontario, L5R 4J3

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging in Publication Data

Bruce, Nigel, 1955–
Quantitative methods for health research : a practical interactive guide / Nigel Bruce, Daniel Pope, and Debbi Stanistreet.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-0-470-02274-0 (cloth : alk. paper) — ISBN 978-0-470-02275-7 (pbk. : alk. paper)
1. Medicine—Research—Methodology. 2. Health—Research—Methodology.
3. Epidemiology—Research—Methodology. 4. Epidemiology.
I. Pope, Daniel, 1969– II. Stanistreet, Debbi, 1963– III. Title.
[DNLM: 1. Epidemiologic Methods. 2. Biomedical Research—methods.
3. Biometry—methods. WA 950 B918q 2008]
RA852.B78 2008
362.1072′4—dc22

2008002734

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 978-0-470-02274-0 (HB) 978-0-470-02275-7 (PB)

Typeset in 10/12pt Times by Integra Software Services Pvt. Ltd, Pondicherry, India
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire
This book is printed on acid-free paper

Contents

Preface

Introduction

Welcome to *Quantitative Methods for Health Research*, a study programme designed to introduce you to the knowledge and skills required to make sense of published health research, and to begin designing and carrying out studies of your own.

The book is based closely on materials developed and tested over almost ten years with the Master of Public Health (MPH) programme at the University of Liverpool, UK. A key theme of our approach to teaching and learning is to ensure a reasonable level of theoretical knowledge (as this helps to provide a solid basis to knowledge), while placing at least as much emphasis on the application of theory to practice (to demonstrate what actually happens when theoretical 'ideals' come up against reality). For these reasons, the learning materials have been designed around a number of published research studies and information sources that address a variety of topics from the UK, Europe and developing countries. The many aspects of study design and analysis illustrated by these studies provide examples which are used to help you understand the fundamental principles of good research, and to practise these techniques yourself.

The MPH programme on which this book is based, consists of two postgraduate taught modules, one **Introductory**, the other **Advanced**, each of which requires 150 hours of study (including assessments), and provides 15 postgraduate credits (1 unit). As students and tutors using the book may find it convenient to follow a similar module-based approach, the content of chapters has been organised to make this as simple as possible. The table summarising the content of each chapter on pages xi to xiii indicates which sections (together with page numbers) relate to the introductory programme, and which to the advanced programme.

The use of computer software for data analysis is a fundamental area of knowledge and skills for the application of epidemiological and statistical methods. A complementary study programme in data analysis using SPSS (Statistical Package for the Social Sciences) has been prepared; this relates closely to the structure and content of the book. Full details of this study programme, including the data sets used for data analysis exercises, are available on the companion web site for this book www.wileyeurope.com/college/bruce.

The book also has a number of other features designed to enhance learning effectiveness, summarised in the following sections.

Learning objectives

Specific, detailed learning objectives are provided at the start of each chapter. These set out the nature and level of knowledge, understanding and skills required to achieve a good standard at Masters level, and can be used as one point of reference for assessing progress.

Resource papers and information sources

All sections of published studies that are absolutely required, in order to follow the text and answer self-assessment exercises, (see below) are reproduced as excerpts in the book. However, we strongly recommend that all resource papers be obtained and read fully, as indicated in the text. This will greatly enhance understanding of how the methods and ideas discussed in the book are applied in practice, and how research papers are prepared. All papers are fully referenced, and the majority are recent and in journals that are easily available through higher education establishments.

Key terms

In order to help identify which concepts and terms are most important, those regarded as core knowledge appear in ***bold italic*** font, thus. These can be used as another form of self-assessment, as a good grasp of the material covered in this book will only have been achieved if all these key terms are familiar and understood.

Self assessment exercises

Each chapter includes self-assessment exercises, which are an integral part of the study programme. These have been designed to assess and consolidate understanding of theoretical concepts and competency in practical techniques. The exercises have also been designed to be worked through as they are encountered, as many of the answers expand on issues that are introduced in the main text. The answers and discussion for these exercises are provided at the end of each chapter.

Mathematical aspects of statistics

Many people find the mathematical aspects of statistical methods, such as formulae and mathematical notation, quite challenging. It is useful however, to gain a basic mathematical understanding of the most commonly used statistical concepts and methods. We recognise that creating the expectation of much more in-depth knowledge for all readers would be very demanding, and arguably unnecessary.

On the other hand, readers with more affinity for and knowledge of mathematics may be interested to know more, and such understanding is important for more advanced research work and data analysis. In order to meet these objectives, all basic concepts, simple mathematical formulae, etc., the understanding of which can be seen as core knowledge are included in the main text. More detailed explanations, including some more complex formulae and examples, are included in statistical reference sections [RS], marked with a start and finish as indicated below.

RS - Reference section on statistical methods

Text, formulae and examples

RS - ends

This book does not set out to provide detailed mathematical explanations of statistical methods as there are plenty of other books that do so and can be referred to if required.

We hope that you will enjoy this study programme, and find that it meets your expectations and needs.

Organisation of subject matter by chapter

The following table summarises the subject content for each chapter, indicating which sections are introductory and which advanced.

Chapter content and level

Introductory (white) **Advanced** (shaded)

Chapter	Level	Pages	Topics covered
1. Philosophy of science and introduction to epidemiology	Introductory	1–28	• Approaches to scientific research • What is epidemiology? • What is Statistics? • Formulating a research question • Rates, incidence and prevalence • Concepts of prevention
2. Routine data sources and descriptive epidemiology	Introductory	29–110	• Routine collection of health information • Descriptive epidemiology • Information on the environment • Displaying, describing and presenting data • Association and correlation • Summary of routinely available data relevant to health • Descriptive epidemiology in action, ecological studies and the ecological fallacy • Overview of epidemiological study designs
3. Standardisation	Introductory	111–128	• Rationale for standardisation • Indirect standardisation • Direct standardisation

Chapter	Level	Pages	Topics covered
4. Surveys	Introductory	129–192	• Rationale for survey methods • Sampling methods • The sampling frame • Sampling error, sample size and confidence intervals • Response rates • Measurement, questionnaire design, validity • Data types and presentation: categorical and continuous
5. Cohort studies	Introductory	193–256	• Rationale for cohort study methods • Obtaining a sample • Measurement and measurement error • Follow-up for mortality and morbidity • Basic analysis – relative risk, hypothesis testing: the *t*-test and the chi-squared test • Introduction to the problem of confounding
	Advanced		• Sample size for cohort studies • Simple linear regression • Multiple linear regression: dealing with confounding factors
6. Case-control studies	Introductory	257–306	• Rationale for case-control study methods • Selecting cases and controls • Matching – to match or not? • The problem of bias • Basic analysis – the odds ratio for unmatched and matched designs
	Advanced		• Sample size for case control studies • Matching with more than one control • Multiple logistic regression
7. Intervention studies	Introductory	307–362	• Rationale for intervention study methods • The randomised controlled trial (RCT) • Randomisation • Blinding, controls and ethical considerations • Basic analysis of trial outcomes: analysis by intention to treat
	Advanced		• Adjustment when confounding factors are not balanced by randomisation • Sample size for intervention studies • Testing more complex health interventions • Factorial design • Cluster randomisation, community randomised trials, quasi-experimental designs, cross-over trials
8. Life tables, survival analysis and Cox regression	Advanced	363–392	• Nature of survival data • Kaplan – Meier survival curves • Cox proportional hazards regression • Introduction to life tables

Chapter	Level	Pages	Topics covered
9. Systematic reviews and meta-analysis	Advanced	393–432	• Purpose of systematic reviews • Method of systematic review • Method of meta-analysis • Special considerations in systematic reviews and meta-analysis of observational studies • The Cochrane Collaboration
10. Prevention strategies and evaluation of screening	Advanced	433–470	• Relative and attributable risk, population attributable risk and attributable fraction • High-risk and population approaches to prevention • Measures and techniques used in the evaluation of screening programmes, including sensitivity, specificity, predictive value and receiver operator characteristic (ROC) curves • Methodological issues and bias in studies of screening programme effectiveness • Cohort and period effects
11. Probability distributions, hypothesis testing and Bayesian methods	Advanced	471–526	• Theoretical probability distributions • Steps in hypothesis testing • Transformation of data • Paired t-test • One-way analysis of variance • Non-parametric tests for paired data, two or more independent groups, and for more than two groups • Spearman's rank correlation • Fisher's exact test • Guide to choosing an appropriate test • Multiple significance testing • Introduction to Bayesian methods

Acknowledgements

The preparation of this book has involved the efforts of a number of people whose support we wish to acknowledge. Francine Watkins for preparing the first section of Chapter 1 'Approaches to Scientific Research'. Jo Reeve for preparing Section 5 of Chapter 2. 'A Summary of Routinely Available Data'. Paul Blackburn for assistance with graphics and obtaining permission for reproduction of the resource papers, and input into the data management and analysis course. Chris West for his advice on statistical methods, and input into the data management and analysis course. Nancy Cleave and Gill Lancaster for their invaluable contributions to early versions of the statistical components of the materials. Our students and programme tutors who have provided much valued, constructive feedback that has helped to guide the development of the materials upon which this book is based. The staff at Wiley for their encouragement and support. Chris and Ian at Ty Cam for providing a tranquil, e-mail free refuge in a beautiful part of Denbighshire.

1

Philosophy of science and introduction to epidemiology

Introduction and learning objectives

In this chapter, we will begin by looking at different approaches to ***scientific research***, how these have arisen, and the importance of recognising that there is no single, 'right way' to carry out investigations in the health field. We will then go on to explore the ***research task***, discuss what is meant by ***epidemiology*** and ***statistics***, and look at how these two disciplines are introduced and developed in the book. The next section introduces the concept of ***rates*** for measuring the frequency of disease or characteristics we are interested in, and in particular the terms ***incidence*** and ***prevalence***. These definitions and uses of rates are fundamental ideas with which you should be familiar before we look in more detail at research methods and study design. In the final section, we will look at key concepts in disease prevention, including the commonly used terms ***primary, secondary*** and ***tertiary*** prevention.

The reason for starting with a brief exploration of the nature of scientific methods is to see how historical and social factors have influenced the biomedical and social research traditions that we take for granted today. This will help you understand your own perceptions of, and assumptions about, health research, based on the knowledge and experience you have gained to date. It will also help you understand the scientific approach being taken in this book, and how this both complements, and differs from, that developed in books and courses on qualitative research methods – as and when you may choose to study these. Being able to draw on a range of research traditions and their associated methods is especially important for the discipline of public health, but also for many other aspects of health and health care.

Learning objectives

By the end of Chapter 1, you should be able to do the following:

Quantitative Methods for Health Research Nigel Bruce, Daniel Pope and Debbi Stanistreet

- Briefly describe the philosophical differences between the main approaches to research that are used in the health field.
- Describe what is meant by epidemiology, and list the main uses to which epidemiological methods and thought can be put.
- Describe what is meant by statistics, and list the main uses to which statistical methods and thought can be put.
- Define and calculate rates, prevalence and incidence, and give examples of their use.
- Define primary, secondary and tertiary prevention and give examples of each.

1.1 Approaches to scientific research

1.1.1 History and nature of scientific research

Scientific research in health has a long history going back to the classical period. There are threads of continuity, as well as new developments in thinking and techniques, which can be traced from the ancient Greeks, through the fall of the Roman Empire, the Dark Ages and the Renaissance, to the present time. At each stage, science has influenced, and been influenced by, the culture and philosophy of the time. Modern scientific methods reflect these varied historical and social influences. So it is useful to begin this brief exploration of scientific health research by reflecting on our own perceptions of science, and how our own views of the world fit with the various ways in which research can be approached. As you read this chapter you might like to think about the following questions:

- What do you understand by the terms ***science***, and ***scientific research***, especially in relation to health?
- How has your understanding of research developed?
- What type of research philosophy best fits your view of the world, and the problems you are most interested in?

Thinking about the answers to these questions will help you understand what we are trying to achieve in this section, and how this can best support the research interests that you have and are likely to develop in the years to come. The history and philosophy of science is of course a whole subject in its own right, and this is of necessity a very brief introduction.

Scientific reasoning and epidemiology

Health research involves many different scientific disciplines, many of which you will be familiar with from previous training and experience. Here we are focusing principally on epidemiology, which is concerned with the study of the distribution and determinants of disease within and between populations. In epidemiology, as we shall see subsequently, there is an emphasis on ***empiricism***, that is, the study of observable phenomena by scientific methods, detailed observation and accurate measurement. The scientific approach to epidemiological investigation has been described as:

- **Systematic** – there is an agreed system for performing observations and measurement.
- **Rigorous** – the agreed system is followed exactly as prescribed.
- **Reproducible** – all the techniques, apparatus and materials used in making the observations and measurements are written down in enough detail to allow another scientist to reproduce the same process.
- **Repeatable** – scientists often repeat their own observations and measurements several times in order to increase the reliability of the data. If similar results are obtained each time, the researcher can be confident the phenomena have been accurately recorded.

These are characteristics of most epidemiological study designs and will be an important part of the planning and implementation of the research. However, this approach is often taken for granted by many investigators in the health field (including epidemiologists) as the only way to conduct research. Later we will look at some of the criticisms of this approach to scientific research but first we need to look in more detail at the reasoning behind this perspective.

Positivism

The assumptions of contemporary epidemiological investigations are associated with a view of science and knowledge known as ***positivism***. Positivism is a philosophy that developed in the eighteenth century in a period known as the Enlightenment, a time when scientists stopped relying on religion, conjecture and faith to explain phenomena, and instead began to use reason and rational thought. This period saw the emergence of the view that it is only by using scientific thinking and practices that we can reveal the truth about the world (Bilton *et al.*, 2002).

Positivism assumes a stable observable reality that can be measured and observed. So, for positivists, scientific knowledge is proven knowledge, and theories are therefore derived in a systematic, rigorous way from observation and experiment. This approach to studying human life is the same approach that scientists take to study the natural world. Human beings are believed by positivists to exist in causal relationships that can be empirically observed, tested and measured (Bilton *et al.*, 2002), and to behave in accordance with various laws. As this reality exists whether we look for it or not, it is the role of scientists to reveal its existence, but not to attempt to understand the inner meanings of these laws or express personal opinions about these laws. One of the primary characteristics of a positivist approach is that the researcher takes an objective distance from the phenomena so that the description of the investigation can be detached and undistorted by emotion or personal bias (Davey, 1994). This means that within epidemiology, various study designs and techniques have been developed to increase objectivity, and you will learn more about these in later chapters.

Induction and deduction

There are two main forms of scientific reasoning – ***induction*** and ***deduction***. Both have been important in the development of scientific knowledge, and it is useful to appreciate the difference between the two in order to understand the approach taken in epidemiology.

Induction

With inductive reasoning, researchers make repeated observations and use this evidence to generate theories to explain what they have observed. For example, if a researcher made a number of observations in different settings of women cooking dinner for their husbands, they might then inductively derive a general theory:

> All women cook dinner for their husbands.

Deduction

Deduction works in the opposite way to induction, starting with a theory (known as an ***hypothesis***) and then testing it by observation. Thus, a very important part of deductive reasoning is the formulation of the hypothesis – that is, the provisional assumption researchers make about the population or phenomena they wish to study before starting with observations. A good hypothesis must enable the researcher to test it through a series of ***empirical observations***. So, in deductive reasoning, the hypothesis would be:

> All women will cook dinner for their husbands.

Observations would then be made in order to test the validity of this statement. This would allow researchers to check the consistency of the hypothesis against their observations, and if necessary the hypothesis can be discarded or refined to accommodate the observed data. So, if they found even one woman not cooking for her husband, the hypothesis would have to be re-examined and modified. This characterises the approach taken in epidemiology and by positivists generally.

One of the most influential science philosophers of recent times was Karl Popper (1902–1994), who argued that hypotheses can never be proved true for all time and scientists should aim to refute their own hypothesis even if this goes against what they believe (Popper, 1959). He called this the ***hypothetico-deductive method***, and in practice this means that an hypothesis should be capable of being falsified and then modified. Thus, to be able to claim the hypothesis is true would mean that all routes of investigation had been carried out. In practice, this is impossible, so research following this method does not set out with the intention of proving that an hypothesis is true. In due course we will see how important this approach is for epidemiology and in the statistical methods used for testing hypotheses.

Alternative approaches to research

It is important to be aware that positivism is only one approach to scientific research. Positivism has been criticised by some researchers, in particular social scientists, who think it is an inappropriate approach to studies of human behaviour. From this perspective, they believe that human beings can behave irrationally and do not always act in accordance with any observable rules or laws. This makes them different from phenomena in the natural world, and so they need to be studied in a different way. Positivism has also been criticised because it does not allow for the view that human beings act in response to others around them; that is, that they interpret their own behaviour in response to others. As Green and Thorogood (2004, p. 12) argue:

> Unlike atoms (or plants or planets), human beings make sense of their place in the world, have views on researchers who are studying them, and behave in ways that are not determined in law-like ways. They are complex, unpredictable, and reflect on their behaviour. Therefore, the methods and aims of the natural sciences are unlikely to be useful for studying people and social behaviour: instead of explaining people and society, research should aim to understand human behaviour.

Social scientists therefore tend to have a different belief about how we should research human beings. Consequently, they are more likely to take an ***inductive*** approach to research because they would argue that they do not want to make assumptions about the social world until they have observed it in and for itself. They, therefore, do not want to formulate hypotheses because they believe these are inappropriate for making sense of human action. Rather, they believe that human action cannot be explained but must be understood.

While positivists are concerned mainly with observing patterns of behaviour, other researchers principally wish to understand human behaviour. This latter group requires a different starting point that will encompass their view of the world, or different ***theoretical positions*** to make sense of the world. It turns out that there are many different positions that can be adopted, and while we cannot go into them all here, we will use one example to illustrate this perspective.

Interpretative approaches

An interpretative approach assumes an interest in the meanings underpinning human action, and the role of the researcher is therefore to unearth that meaning. The researcher would not look to measure the reality of the world but would seek to understand how people interpret the world around them (Green and Thorogood, 2004).

Let's look at an example of this in respect of asthma. A ***positivist*** approach to researching this condition may be to obtain a series of objective measurements of symptoms and lung function by a standard procedure on a particular sample of people over a specified period of time. An ***interpretative*** approach might involve talking in-depth to fewer participants to try to understand how their symptoms affect their lives. Obviously, in order to do this, these two 'types' of researchers would need to use very different approaches. Those planning the interpretative research would be more likely to use ***qualitative methods*** (interviews, focus groups, participatory methods, etc.), while positivists (for example, epidemiologists) would choose ***quantitative methods*** (surveys, cohort studies, etc., involving lung-function measurements and highly structured questionnaires). These two different approaches are called ***research paradigms*** and would therefore produce different types of information.

Interpretative researchers would also criticise positivists for their belief that researchers can have an objective, unimpaired and unprejudiced stance in the research that allows them to make value-free statements. Interpretative researchers accept that researchers are human beings and therefore cannot stand objectively apart from the research. In a sense they are part of the research, as their presence can influence the nature and outcome of the research. Whether or not you agree with the criticism of positivism, you need to be aware that there are alternative approaches to conducting research that neither prioritise objectivity nor set out to measure 'reality'.

One of the most influential writers on the scientific method was Thomas Kuhn (1922–1996) (Davey, 1994). He argued that one scientific paradigm – one 'conceptual worldview' – may be the dominant one at a particular period in history. Over time, this is challenged, and eventually replaced by another view (paradigm), which then becomes accepted as the most important and influential. These revolutions in science were termed 'paradigm shifts'. Although challenged by other writers, this perspective suggests that scientific methods we may take for granted as being the only or best way to investigate health and disease, are to an extent the product of historical and social factors, and can be expected to evolve – and maybe change substantively – over time.

Exercise for reflection

1. Make brief notes on the type of scientific knowledge and research with which you are most familiar.
2. Is this predominantly positivistic (hypothetico-deductive) or interpretative in nature, or is it more of a mixture?

There are no 'answers' provided for this exercise, as it is intended for personal reflection.

1.1.2 What is epidemiology?

The term *epidemiology* is derived from the following three Greek words:

Epi **– among**
Demos **– the people**
Logos **– discourse**

We can translate this in more modern terms into ***'The study of the distribution and determinants of disease frequency in human populations'***. The following exercise will help you to think about the uses to which the discipline of epidemiology is put.

Self-Assessment Exercise 1.1.1

Make a list of some of the ***applications*** of epidemiological methods and thought that you can think of. In answering this, avoid listing types of epidemiological study that you may already know of. Try instead to think in general terms about the practical outcomes and applications of these methods.

Answers in Section 1.5

This exercise shows the very wide application of epidemiological methods and thought. It is useful to distinguish between two broad functions of epidemiology, one very practical, the other more philosophical:

- The range of epidemiological research methods provides a toolbox for obtaining the best scientific information in a given situation (assuming, that is, you have established that a positivist approach is most appropriate for the topic under study!).
- Epidemiology helps us use knowledge about the population determinants of health and disease to inform the full range of investigative work, from the choice of research methods, through analysis and interpretation, to the application of findings in policy. With experience, this becomes a way of thinking about health issues, over and above the mere application of good methodology.

You will find that your understanding of this second point grows as you learn about epidemiological methods and their application. This is so because epidemiology provides the means of describing the characteristics of populations, comparing them, and analysing and interpreting the differences, as well as the many social, economic, environmental, behavioural, ecological and genetic factors that determine those differences.

1.1.3 What are statistics?

A statistic is a numerical fact. Your height and weight and the average daily rainfall in Liverpool are examples of statistics. The academic discipline of statistics is concerned with the collection, presentation, analysis and interpretation of numerical information (also called ***quantitative*** information).

Statistics are everywhere!

We are surrounded by, and constantly bombarded with, information from many sources - from the cereal box, to unemployment figures, the football results and opinion polls, to articles in scientific journals. The science of statistics allows us to make sense of this information and is thus a fundamental tool for investigation in many disciplines, including health, education, economics, agriculture and politics, to name but a few. The next exercise encourages you to explore how statistics are used in everyday life.

Self-Assessment Exercise 1.1.2

Look at a recent newspaper (hard copy or website for a newspaper, or, for example, the BBC news website (http://news.bbc.co.uk)) and find up to five items in which statistics are used. List the ways in which numerical information is presented.

Examples in Section 1.5

The scientific term for pieces of information is ***data***. The singular is ***datum***, meaning a single piece of information, such as, for example, one person's weight. A set of data may consist of many items, such as the heights, weights, blood pressures, smoking habits and exercise level of several hundred people. In its raw state, this mass of figures tells us little. There are two ways in which we use statistics to help us interpret data:

- To ***describe*** the group about which the data have been collected. This may be a group of people, or a group of hospitals, or a group of laboratory cultures. We describe the group by summarising the information into a few meaningful numbers and pictures.
- To ***infer*** something about the population of which the group studied is a part. We often want to know something about a population, such as everyone aged over 65 in Liverpool, but, practically, can collect information about only a subset of that population. This 'subset' is called a sample, and is explored in Chapter 3 on surveys. With inference, we want to know what generalisations to the population can be made from the sample, and with what degree of certainty.

Self-Assessment Exercise 1.1.3

Can you find one example of ***description***, and one example of ***inference*** in your newspaper or Web search? If you have found an example of making an inference, to which population does it apply?

Examples in Section 1.5

1.1.4 Approach to learning

We will explore the use and interpretation of statistical techniques through a number of published studies. Whether or not you go on to carry out research and use statistical methods yourself, you are certain to be a consumer of statistics through published research. We shall emphasise both the use of appropriate techniques and the critical interpretation of published results. You will also be learning about epidemiology and statistics in an integrated way. This approach recognises that the two disciplines embody many closely related concepts and techniques. There are also certain very distinct qualities, which you will find are emphasised through the more theoretical discussion relating to one or other discipline. Your learning of these research methods is based primarily on practical examples of data and published studies in order to help you to see how epidemiology and statistics are used in practice, and not just in theoretical or ideal circumstances.

Summary

- There is no single, 'right' philosophy of research. The approach taken is determined by many factors, including historical and social influences, and the nature of the problem being investigated.
- As an individual, your education, training and experience will strongly influence the scientific 'paradigm' that you are familiar, and comfortable with.
- A variety of approaches to, and methods for, research is both appropriate and necessary in the health field.
- Epidemiology provides us with a range of research tools, which can be used to obtain the information required for prevention, service provision, and the evaluation of health care. One of the most important contributions of epidemiology is the insight gained about the factors which determine the health of populations.
- Statistics is concerned with the collection, presentation, analysis and interpretation of numerical information. We may use statistical techniques to describe a group (of people, hospitals, etc.) and to make inferences about the population to which the group belongs.

1.2 Formulating a research question

1.2.1 Importance of a well-defined research question

This is arguably the most important section of the whole book. The reason for our saying this is that the methods we use, and ultimately the results that we obtain, must be determined by the question we are seeking to answer.

So, how do we go about formulating that question? This does not (usually) happen instantly, and there is a good deal of debate about how the question ought to be formulated, and how it is formulated in practice. Karl Popper argued that research ideas can come from all kinds of sources. But the idea is not enough on its own, and will usually need working on before it is a clearly formulated ***research question***. Here are some of the factors we will have to take into account in fashioning a clear research question:

- What does all the other work on the topic tell us about the state of knowledge, and what aspects need to be addressed next? Our idea might actually arise from a review such as this, and therefore be already fairly well defined, but more often than not, the idea or need arises before we have had a chance to fully evaluate the existing body of knowledge. This will also depend on how far we are into researching a particular subject.
- Different types of problem and topic areas demand, and/or have been traditionally associated with, different research traditions. Does our idea require a positivist approach, with an hypothesis that can be falsified? If so, the research question will need to be phrased in a way that allows this. Alternatively, we might be trying to understand how a certain group of people view a disease and the services provided for them. This question must also be precisely defined, but not in the same way: there is nothing here to be falsified; rather, we wish to gain as full an understanding as possible of people's experience and opinions to guide the development of services.

- If our idea is overambitious, the necessary research may be too demanding, expensive or complex to answer the question(s) in one go. Perhaps it needs to be done in separate studies, or in stages.

In practice, defining the research question does not usually happen cleanly and quickly, but is a process that gradually results in a more and more sharply defined question as the existing knowledge, research options, and other practical considerations are explored and debated. There may appear to be exceptions to this – for instance, a trial of a new drug. On the face of it, the question seems simple enough: the drug is now available, so is it better than existing alternatives or not? However, as we will discover later, the context in which the drug would be used can raise a lot of issues that will play a part in defining the research question.

Although knowledge of appropriate research methods is important in helping you to formulate a clear research question, it is nevertheless useful to start the ***process*** of developing your awareness of, and skills in, this all-important aspect of research. The following exercise provides an opportunity for you to try this.

Self-Assessment Exercise 1.2.1

A research idea . . .

> Your work in an urban area involves aspects of the care and management of people with asthma. You are well aware of the contemporary concern about the effect of air pollution on asthma, and the view that while pollution (e.g. ozone, nitrogen oxides) almost certainly exacerbates asthma, this may not be the cause of the underlying asthmatic tendency.
>
> In recent years, you have noticed that asthmatics (especially children) living in the poorest parts of the city seem to suffer more severe and frequent asthma episodes than those living in better-off parts.
>
> Although you recognise that pollution from traffic and industry might be worse in the poorer areas, you have been wondering whether other factors, such as diet (e.g. highly processed foods) or housing conditions such as dampness and associated moulds might be the real cause of the difference.
>
> You have reviewed the literature on this topic and found a few studies which are somewhat conflicting, and do not seem to have distinguished very well between the pollution, diet and housing factors you are interested in.

Have a go at converting the idea described above into a well-formulated research question appropriate for epidemiological enquiry. Note that there is no single, right, research question here. You do not need to describe the study methods you might go on to use.

Specimen answer in Section 1.5

1.2.2 Development of research ideas

We have seen that research is a process which evolves over a period of time. It is influenced by many factors, including other work in the field, the prevalent scientific view, political factors, finance, etc. Research is not a socially isolated activity with a discrete (inspirational) beginning, (perfect) middle and (always happy) ending, Figure 1.2.1.

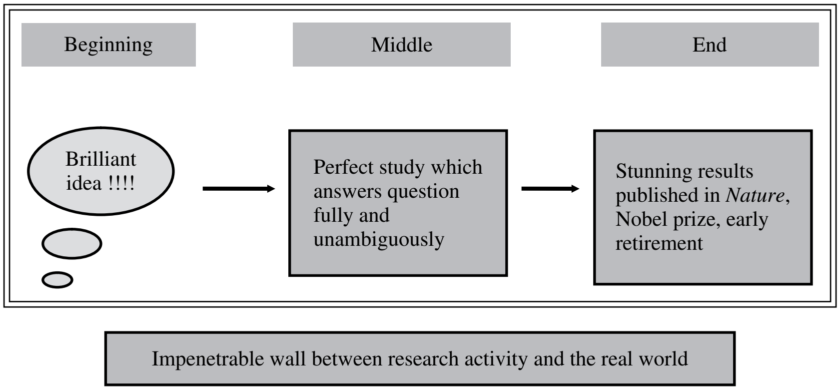

Figure 1.2.1 Research fantasy time

A more realistic way to describe the process of research development is cyclical, as illustrated in Figure 1.2.2. A well-defined and realistic question, which is (as we have seen) influenced by many factors, leads to a study which, it is hoped, provides much of the information required.

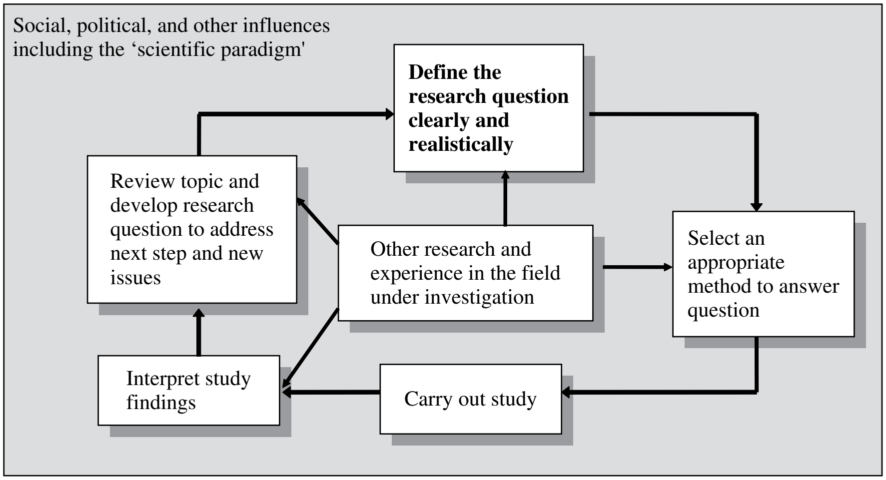

Figure 1.2.2 Research is a process that can usefully be thought of as being cyclical in nature, and subject to many influences from both inside and outside the scientific community

These findings, together with developments in the scientific field as well as social, political or other significant influences, will lead to further development of the research question.

Summary

- Defining a clear research question is a fundamental step in research.
- Well-defined research questions do not (usually) appear instantly!
- Research is a process, which can usefully be thought of as cyclical, albeit subject to many external influences along the way.

1.3 Rates: incidence and prevalence

1.3.1 Why do we need rates?

A useful way to approach this question is by considering the problem that arises when we try to interpret a change in the number of events (which could be deaths, hospital admissions, etc.) occurring during, say, a period of one year in a given setting. Exercise 1.3.1 is an example of this type of problem, and is concerned with an increase in numbers of hospital admissions.

Self-Assessment Exercise 1.3.1

Over a period of 12 months, the accident and emergency department of a city hospital noted the number of acute medical admissions for people over 65 had increased by 30 per cent. In the previous 5 years, there had been a steady increase of only about 5 per cent per year.

1. List the possible reasons for the 30 per cent increase in hospital accident and emergency admissions.
2. What other information could help us interpret the reasons for this sudden increase in admissions?

Answers in Section 1.5

In this exercise we have seen the importance of interpreting changes in numbers of events in the light of knowledge about the ***population*** from which those events arose. This is why we need ***rates***. A rate has a ***numerator*** and a ***denominator***, and must be determined over a specified ***period of time***. It can be defined as follows:

"Numerator"

$$\text{RATE} = \frac{\textbf{Number of events arising from defined population in a given period}}{\textbf{Number in defined population, in same period}}$$

"Denominator"

1.3.2 Measures of disease frequency

We will encounter rates on many occasions, as these are a fundamental means of expressing information gathered and analysed in epidemiological research. We can view these as measures of the ***frequency*** of disease, or of characteristics in the population. Two important general ways of viewing disease frequency are provided by ***prevalence*** and ***incidence***.

1.3.3 Prevalence

The prevalence tells us how many cases of a disease (or people with a characteristic, such as smoking) there are in a given population, at a specified time. The ***numerator*** is the number of cases, and the ***denominator*** the population we are interested in.

$$\textbf{\textit{Prevalence}} = \frac{\text{Number of cases at a given time}}{\text{Number in population at that time}}$$

This can be expressed as a percentage, or per 1000 population, or per 10 000, etc., as convenient. The following are therefore examples of prevalence:

- In a primary care trust (PCT), with a population of 400 000, there are 100 000 smokers. The prevalence is therefore 25 per cent, or 250 per 1000.
- In the same PCT, there are known to be 5000 people diagnosed with schizophrenia. The prevalence is therefore 1.25 per cent, or 12.5 per 1000.

Note that these two examples represent 'snapshots' of the situation at a given time. We do not have to ask about people starting or giving up smoking, nor people becoming ill with (or recovering from) schizophrenia. It is a matter of asking, 'in this population, how many are there now?' This snapshot approach to measuring prevalence is known as ***point prevalence***, since it refers to one point in time, and is the usual way in which the term prevalence is used. If, on the other hand, we assess prevalence over a period of time, it is necessary to think about cases that exist at the start of the period, and new cases which develop during the period. This measure is known as ***period prevalence***, and this, together with point prevalence, is illustrated in Figure 1.2.3 and Exercise 1.3.2.

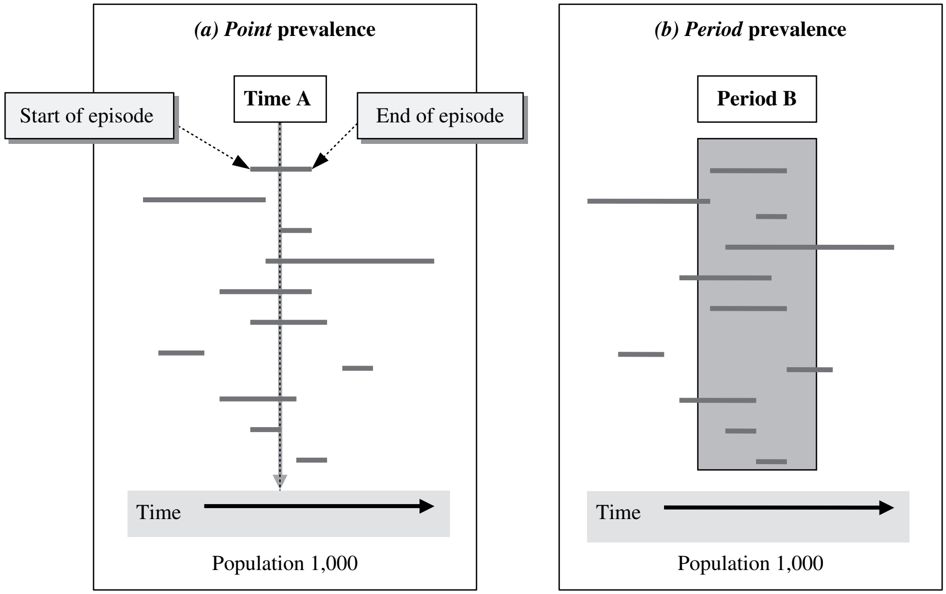

Figure 1.2.3 Period and point prevalence

Point prevalence (Figure 1.2.3a) is assessed at one point in time (time A), whereas period prevalence (Figure 1.2.3b) is assessed over a period (period B). The horizontal bars represent patients becoming ill and recovering after varying periods of time – the start and end of one episode is marked in Figure 1.2.3a. Period prevalence includes everyone who has experienced the disease at some time during this period.

Self-Assessment Exercise 1.3.2

1. In the above examples (Figure 1.2.3a and b), calculate the point prevalence and period prevalence.
2. Why do the point prevalence and period prevalence differ?

Answers in Section 1.5

1.3.4 Incidence

Whereas ***prevalence*** gives us a measure of how many cases there are in the population at a given time, ***incidence*** tells us the rate at which ***new*** cases are appearing in the population. This is a vital distinction. In order to determine the incidence, we need to know the number of new cases appearing over a specified period of time, and the number of people in the population who could