Part I

Introduction

1

Scope of epidemiological enquiry and overview of main problem areas

1.1 What questions can epidemiology answer?

Epidemiology can be defined as the study of the distribution of disease and its determinants in human populations. In other words, it provides the answers to questions on how much disease there is, who gets it and what specific factors put individuals at risk. The epidemiology section in a medical textbook chapter on a particular disease will provide data on these aspects. There is an alternative and broader view of epidemiology, which is that it is a methodology to obtain answers about diseases from their study in humans. This broader definition allows a substantially greater scope for the kinds of question that can be addressed both by those studying the health of populations and by those whose main focus is the study of disease in patient groups. The list in Table 1.1 represents the vast array of topics that epidemiologists would consider as relevant to their discipline.

1.1a Disease definition

Most diseases lack a clear diagnostic test that totally discriminates between disease and normality, though infectious disease and trauma are two obvious exceptions. Most often the diagnosis is based on clinical opinion, with the latter based on experience, prejudice or arbitrary rules. In the absence of a standardised definition of disease, results from aetiological, prognostic or therapeutic studies cannot be directly compared. The development of disease criteria is a separate topic in itself which requires a careful epidemiological study of the occurrence of specific features in cases determined by a notional gold standard, such as the expert clinician's diagnosis. These features are then compared with those from an appropriate group of non-cases and the level of agreement evaluated.

Table 1.1. Questions relevant for epidemiological enquiry

Disease definition	What characteristics or combination of characteristics best discriminate disease from non-disease?
Disease occurrence	What is the rate of development of new cases in a population? What is the proportion of current disease within a population? What are the influences of age, sex, time and geography on the above?
Disease causation	What are the risk factors for disease development and what are their relative strengths with respect to an individual and a population?
Disease outcome	What is the outcome following disease onset and what are the risk factors, including their relative strengths, for a poor outcome?
Disease management	What is the relative effectiveness of proposed therapeutic interventions? (Included within this are health service research questions related to the relative effectiveness of proposed health service delivery options.)
Disease prevention	What is the relative effectiveness of proposed preventive strategies including screening?

Example 1.i

Prior to starting a research programme on migrainous headaches, a neurologist realised that it was necessary to derive criteria that (i) were easy to apply, (ii) could distinguish migraine from other causes of headaches and (iii) would be accepted by the neurological community.

1.1b Disease occurrence

This is the classical focus of epidemiology and the available approaches to measure occurrence are discussed in Chapter 2. Data on occurrence are of interest in their own right, but are also relevant both to the clinician, in weighing up different diagnostic likelihoods in the face of the same evidence, and to those providing health services. A more detailed study will uncover differences in occurrence between sexes and across age groups, over time and between different geographical populations. Indeed, the age and sex effects on disease occurrence are normally so strong that it is absolutely fundamental to gather such data in order to compare disease occurrence both between

populations and within the same population over time. These issues are discussed in Chapter 3. In addition, marked differences between occurrence in different population groups may provide aetiological clues for further enquiry.

1.1c Disease causation

Similarly, the use of epidemiology to unravel causative mechanisms is one of its major roles. It is, however, too simplistic for most chronic diseases, to consider their influence on disease risk as present or absent. It is the strength of any disease association with possible risk factor variables that is of more interest.

Example 1.ii

In planning a study on whether workers exposed to organic dusts were at increased risk of various lung diseases, the investigators aimed to discover (i) whether or not there was an increased risk, (ii) the level of any increase for the major lung disorders considered and (iii) how these risks compared with those from smoking in the same population.

Risk and association

It is appropriate, at this stage, to clarify the meaning of the terms 'risk' and 'association'. In common use, association indicates a relationship with no indication as to the underlying path. As an example, there is an association between an individual's height and his/her weight, although there are a number of possible paths: (i) the taller an individual the greater will be the weight; (ii) (unlikely) the heavier an individual the greater will be the height; or (iii) there are common factors, for example genetic, that influence both height and weight. By contrast, risk implies that the pathway is known (or worthy of investigation). Thus in the example above, the question can be addressed whether height is a risk factor for (being over) weight. In practice, epidemiological investigations aim to uncover associations that, using other information, are translated into risks.

1.1d Disease outcome

Investigations concerning the frequency and prediction of specific disease outcomes in patient populations may be considered as the clinical epidemiological parallels of studies of disease occurrence and causation in normal

populations. Thus the *population* epidemiologist may wish to ascertain the incidence of, and risk factors for, angina in a stated population; whereas the *clinical* epidemiologist may wish to ascertain the incidence of, and risk factors for, subsequent myocardial infarction and sudden death in patients with angina. The methodological concepts, however, are identical, as will be discussed in later chapters.

1.1e Disease management and disease prevention

The use of the clinical trial to evaluate the effectiveness of a particular therapeutic intervention is well established in medicine. Epidemiologically, the clinical trial can be considered as an experimental study where the investigator has intervened to alter the 'exposure', e.g. management, in order to measure the effect on disease occurrence or outcome. The term 'intervention study' describes this broad activity. A further aim of this type of study is to determine whether a link between a suspected risk factor and disease is causative rather than simply an association.

Example 1.iii

In order to examine the possibility that dietary folate deficiency during pregnancy was a causative factor for neural tube defects, an intervention study was carried out on high-risk pregnant women who were randomly allocated to receive folate supplementation or placebo.

Conversely, intervention trial concepts can be applied to health service delivery to answer questions such as whether policy A is likely to be more effective than policy B in reducing waiting lists. Health service research questions such as this require the same rigorous epidemiological approach. In most developed countries with increasing economic pressure to contain health service costs, there is considerable demand (and funding) for epidemiologists to apply their expertise in this area.

An extension of the above is the use of the intervention trial to assess the effectiveness of a population-wide preventive strategy. Population-based primary prevention trials can indeed be considered as clinical trials on a massive scale. Screening for preclinical disease is a widely practised preventive strategy and the evaluation of screening as a tool for reducing morbidity/mortality can be considered under the same heading as above.

1.2 What are the major issues in conducting epidemiological research?

Much of the above seems straightforward, and indeed part of the attraction of epidemiology is its accessibility to the potential investigator. Compared with other approaches to studying biomedical issues, epidemiology often does not require expensive or highly technical equipment and superficially, at least, its language and concepts are those of everyday 'medical speak' that do not require the initiation into a new language as does molecular biology or immunology, for example. There are, however, distinct epidemiological concerns, both for the first-time investigator and for the expert reviewing the work of others, stemming in a large part from the basic tenet that epidemiology deals with 'free living', and frequently healthy, human populations. The consequences of this are: (i) methods of study have to be simple and non-invasive; (ii) subjects, as compared with laboratory animals, may choose to participate or not, or even withdraw from participation during a study; and (iii) the experimental approach where the investigator modifies the conditions to study a specific factor is fraught with difficulties and, as a result, experimental studies are infrequent in epidemiological research. In addition, since many important diseases are relatively rare, studies often need to be large in scope, long in duration or both, with consequences both for the resources required and for the patience and longevity of the investigator.

There are a substantial number of problems to be considered in undertaking any epidemiological study. These are listed in Table 1.2, which provides the framework for the rest of the volume, and are discussed in outline below.

1.2a Study design

The first demand is to clearly frame and thereafter focus on the specific questions posed. In the following two chapters, the various options for studies of disease occurrence and causation are outlined. A decision has to be made about the choice of study design that can best answer the question posed, taking into account the often conflicting demands of scientific validity and practicality.

1.2b Population selection

The subjects to be studied have to be defined both in terms of the group(s) from which they are to be selected and, in selecting individuals, the inclusion

Table 1.2. Major problem areas for epidemiological research

Study design	What is the question posed – what type of study can best answer the question and is most practicable?
Population selection	Who should be studied? How many should be studied?
Information gathering	How should the information be obtained? Is the information obtained correct? Is the method used to obtain the information consistent?
Analysis	How should the data gathered be prepared for analysis? What are the appropriate analytical methods?
Interpretation of results	Can any associations observed be explained by confounding? Are the results explained by bias? Are the results generalisable?
Logistics	Is the research ethical? Is the research affordable?

and exclusion rules. Specific problems arise in comparative studies when it is necessary to recruit two or more groups based on their disease or on their risk factor status. Problems in population selection are one of the major reasons for a study's conclusions being invalid. A specific difficulty is that of sample size. Cost, time, or other practical considerations may limit the availability of subjects for study. A scientific approach to sample size estimation is given for the different study design options later on in the book. Non-response or loss to follow-up can reduce the number of subjects available for analysis and an adequately large study at the onset may prove too small by the end.

1.2c Information quality

This major issue relates to the quality of the data obtained. There is a particular problem when the approach requires a subject to recall past symptoms or exposures. The most appropriate method for obtaining information must be selected. This might, for example, be a choice between interview and self-administered questionnaire. Other sources of information such as data collected prior to the study, often for another purpose such as the medical record, may be available. The classical approach is to consider the quality of

information obtained under the headings of: (i) *validity*, i.e. does the measurement give the true answer, and (ii) *repeatability*, i.e. is the same answer obtained from the same person when repeated measures are made?

1.2d Data handling and analysis

The time spent on this activity is frequently longer than that spent on the data collection itself. In particular, there is a need to ensure that the data analysed are complete and error-free. The next problem is to choose the appropriate method of analysis.

1.2e Interpreting the results

The first issue is that of confounding. Put simply, it is often difficult in human populations to distinguish between attributes that frequently occur together. Thus, in studies to determine the effect of cigarette smoking on the risk for a particular disease, a positive association may be observed that does not relate to the direct impact of smoking on risk but reflects the joint association between a *confounder*, such as alcohol consumption, which is linked to both cigarette smoking and the disease under study. One of the major advances in the practice of epidemiology in the past decade has been the simultaneous development of user-friendly software and accessible hardware that permit the analysis of the impact of any potential confounder in a way that manual methods of statistical analysis could not achieve.

The second issue is whether the results obtained could be explained by bias, either in the selection of subjects, in the subjects who chose to participate, or in the gathering of information.

The third issue is whether the results are generalisable. A study has been conducted amongst university students examining the relationship between coffee consumption and migraine headaches. Students with migraine were more than twice as likely to consume, on average, more than two cups of coffee per day. Is the association generalisable, outside the study population?

1.2f Logistical issues

Two important areas to be addressed are those of ethics and cost. Studying free-living individuals imposes ethical constraints on the investigators, and the need for cost containment is self-evident. Indeed, these issues have to be considered early as they are likely to influence the final study design.