MEDICAL PRACTICE

*Epidemiology for the Uninitiated***What is epidemiology?**

GEOFFREY ROSE, D J P BARKER

British Medical Journal, 1978, 2, 803-804

Epidemiology is the study of disease in relation to populations. Like the clinical findings and pathology, the epidemiology of a disease is an integral part of its basic description. The subject has its special techniques of data collection and interpretation, and its necessary jargon of technical terms. It will be the aim of this series of short articles to provide for the uninitiated an ABC of the epidemiological approach, its terminology, and its methods. Our only assumption will be that readers already believe that epidemiological questions are worth answering. This introductory article will indicate some of the distinctive characteristics of the epidemiological approach, especially those which differ from the familiar clinical descriptive methods.

All findings must relate to a defined population

Whereas clinical studies are carried out on patients, epidemiology, which seeks to relate disease to the population in which it occurs, requires study of both sick and healthy. If findings in the "cases" cannot be related to the defined population in which those cases arose, then it is not epidemiology.

This *population at risk* is a basic concept of epidemiology. In a study of accidents to patients in hospital it was noted that the largest number occurred among the elderly, and from this the authors concluded that "patients aged 60 and over are most prone to accidents." Another recent paper, based on a survey

of hang-gliding accidents, recommended that flying should be banned between 11 am and 3 pm, because this was the time when 73% of the accidents occurred. Each of these studies based conclusions on the same logical error—namely, the *floating numerator*: the number of cases was not related to the appropriate "at risk" population. Had this been done, the conclusions might have been different. Epidemiological conclusions (on risk) cannot be drawn from purely clinical data (on the frequency of sick people seen).

What is a population? Statisticians, demographers, and epidemiologists would answer differently. Epidemiologically it means a defined group, including both sick and healthy, about whose health some statement is to be made. In practice, it is unlikely that the whole of this *target population* can or needs to be examined, access being limited to some subset of it—the *study population*. Thus a chemical pathologist wishing to establish a range of normal values may investigate the employees of a local factory, hoping that they represent the generality of healthy people. The process of inference (generalisation) from study to target population is informal, not statistical.

Usually even the study population will be larger than the investigation requires, and actual examination may be limited to a *study sample*. Thus the people examined are now at two removes from the group with which the study is ultimately concerned:

Target population → *study population* → *study sample*

The study sample is of interest only in so far as it shows what is happening in the population. Inference from sample to study population is rigorous and statistical, provided that the sample is representative. This requires (a) an unambiguous *definition* of who is in the study population; (b) an enumeration or *census* of all its members; (c) a *sampling* process such that each member has a known (usually equal) probability of selection into the study sample. Probability sampling does not guarantee representativeness, but it does mean that the limits of sampling error may be precisely calculated.

The definition of the study population begins with some basic characteristic which all its members have in common.

Department of Medical Statistics and Epidemiology, London School of Hygiene and Tropical Medicine, London WC1E 7HT
GEOFFREY ROSE, DM, FRCP, professor of epidemiology

Faculty of Medicine, University of Southampton, Southampton General Hospital, Southampton SO9 4XY
D J P BARKER, PHD, MRCP, reader in clinical epidemiology

This may be *geographical* ("all UK residents in 1978," or "all residents in a certain health district"); *occupational* ("all employees of a factory," "children attending a certain primary school," "all instrument and watchmakers and precision tool operators in England and Wales"*); or *based on special care* ("a GP's list," "residents in a geriatric home," "all persons coming to necropsy").

Within this broad definition appropriate restrictions may be specified—for example, in age range or sex.

Oriented to groups rather than individuals

Clinical observations determine decisions about individuals. Epidemiological observations determine decisions about groups, being the basis for preventive policy and allocation of resources.

This fundamental difference in the purpose of measurements implies different demands on the quality of data. An inquiry into the validity of death certificates as an indicator of the frequency of oesophageal cancer produced the following results:

Diagnosed by clinician	74
Confirmed by pathologist	53
Not confirmed by pathologist	21
First diagnosed post mortem	22

Inaccuracy was alarming at individual patient level. Nevertheless, the false-positive errors balanced the false-negatives, and so the clinicians' total (74 cases) was about the same as the pathologists' total (53 + 22 = 75 cases). Hence mortality statistics in the population seem to be about right, despite the unreliability of any individual death certificate.

Conversely, it may not be too serious clinically if "Doctor X" systematically records blood pressure 10 mm Hg higher than his colleagues, because his management policy is (one hopes) adjusted accordingly. But in a community study of the frequency of hypertension choosing "Doctor X" as observer would be unfortunate.

*A group recently reported to have excessive mortality from arthritis and anaemia, possibly due to potent treatments given because their work demands supple fingers.

Does the appearance of the throat give any clue to whether acute infections of the tonsils and pharynx are caused by a virus or by streptococci? Are there any other reliable clinical indicators as to viral or bacterial cause, and does an associated rhinorrhoea or stuffiness of the nose denote a viral origin?

Acute infections of the tonsils and pharynx may be caused by either viruses or bacteria. Commonly, bacterial infections are superimposed on a pre-existing, diffuse viral inflammation. Reliable indicators of bacterial aetiology are: (1) membrane on the tonsils associated with intense faucial injection and in the absence of rhinorrhoea; (2) strawberry tongue, indicating streptococcal infection; and (3) the membrane that causes bleeding when removed, associated with the characteristic sweet odour of diphtheria. Rhinorrhoea and nasal stuffiness are symptoms of nasal or sinus inflammation, or both. In themselves they offer no indication about the causative agent. Secondary bacterial sinusitis in its early stages presents these appearances, although the symptomatic effects are usually more florid.

A 70-year-old woman has persistent oral pain and glossitis as a result of stones in her salivary glands. So far the glands have not been obstructed. What treatment is advised?

Unfortunately, there is no treatment apart from surgery that can deal competently with stones in the salivary glands. By far the commonest site is the submandibular salivary gland. If the stone is in its duct, it can simply be removed through the floor of the mouth. Stones in the gland itself, however, call for excision of the gland. This is a

Conclusions are based on comparisons

Clues to *aetiology* come from comparing disease rates in groups with differing levels of exposure—for example, increase in congenital defects after a rubella epidemic, excess lung cancer and pleural mesothelioma in those exposed to asbestos. Clues will be missed, or false clues created, if comparisons are biased by unequal ascertainment of cases or exposure levels. (Or course, if everyone is equally exposed there will not be any clues—epidemiology thrives on heterogeneity.)

Identifying *high-risk* and *priority groups* again rests on unbiased comparison of rates. From the *Decennial Occupational Supplement of the Registrar General of England and Wales* (1970-72) it appears that there are major differences between occupations in the proportion of men surviving to age 65:

Farmers (self-employed)	82%
Professionals	77%
Skilled manual workers	69%
Labourers	63%
Armed Forces	42%

These differences look important and challenging. Firstly, however, one must consider how far the comparison may have been biased in ascertainment of either the deaths or the populations at risk, or by selective influences on recruitment or retirement.

Another task of epidemiology is *monitoring* or *surveillance* of time trends to show which diseases are increasing, which decreasing, and which changing in their distribution. This information is needed for the identification of both emerging problems and of the effectiveness of measures to control old ones. Unfortunately, standards of diagnosis and data recording may change, and conclusions from time trends call for particular wariness.

The data from which epidemiology seeks to draw conclusions are nearly always collected by several people, often from different countries. Rigorous *standardisation* and *quality control* of investigative methods demand major emphasis in epidemiology; and if an apparent difference in disease rates has emerged, the first question is always "Might the comparison be biased?"

Eventually this series will be collected into a book and hence no reprints will be available from the authors.

straightforward operation, and the results are satisfactory. Stones in the parotid gland and its duct are much less common but again, if they cause symptoms, need to be surgically removed. Salivary calculi in this age group are relatively uncommon and are much more often seen in young adults. Interestingly enough, they nearly always occur in patients with an excellent set of teeth and not, as one might imagine, in those with gross dental neglect and caries.

What virus infections of the genital tract are known, and do any of them cause a progressive loss in erectile capability?

The principal viruses affecting the genital tract are herpes virus hominis type 2, the wart virus, and the virus of molluscum contagiosum. In addition, several viruses are now known to be sexually transmissible, and these include the hepatitis B virus, cytomegalovirus, the Epstein-Barr virus, and the Marburg virus, none of which causes progressive loss of erectile capability.

Correction

Use of antibiotics: Endocarditis

In the Today's Treatment article by Dr Celia Oakley (12 August 1978, p 489) the recommended peak and trough blood concentrations of gentamicin should have read "peaks should not exceed 14 mg/l nor be less than 5 mg/l."