Epidemiology for the Uninitiated

Observer variation

GEOFFREY ROSE, D J P BARKER

British Medical Journal, 1978, 2, 1006-1007

Observational data derived from whole populations are crude and often inaccurate, and comparisons when different observers are used are particularly fraught with danger. The epidemiologist has to apply intellectual rigour to interpreting weak data, and his motto could well be "dirty hands, but a clean mind" (*manus sordidae, mens pura*). Observer variation is also important in clinical practice, but the demands of epidemiology have given a special fillip to its study. It is primarily the basis for assessing the quality of the data, and then for improving that quality.

Collecting the information

Every epidemiological study should include for its main measurements some check of their repeatability (also known, less satisfactorily, as "precision" and "reliability"). This is best built in to the main study—either a sample of subjects undergoing a second examination, or a sample of x-ray films, blood samples, and so on being tested in duplicate. Even a small sample is valuable, provided that (1) it is representative, and (2) the duplicate tests are genuinely independent, the observer being unable to identify the pairs. If testing is done "off-line" (perhaps as part of a pilot study), then particular care is needed to ensure that subjects, observers, and operating conditions are all adequately representative of the main study. It is much easier to test repeatability when the data can be transported and stored—for example, deep-frozen plasma samples, histological sections, and all kinds of tracings and photographs. Nevertheless, such tests may exclude an important source of observer variation—namely, the techniques of obtaining samples and records.

Identifying the components of error

Independent replicate measurements in the same subjects are usually found to vary more than one's gloomiest expectations. To interpret the results, and to seek remedies, it is helpful to dissect the total variability into its four components:

Within-observer variation—Discovering one's own inconsistency can be traumatic; it highlights a lack of clear criteria of measurement

and interpretation, particularly in dealing with the grey area between "normal" and "abnormal." It is largely *random*—that is, unpredictable in direction.

Between-observer variation—This includes the first component (the instability of individual observers), but adds to it an extra and *systematic* component due to individual differences in technique and criteria. Unfortunately, this may be large in relation to the real difference between the groups which it was hoped to identify.

It may be possible to avoid the problem altogether, either by using a single observer or, if data are transportable, by forwarding them all for central examination. Alternatively, the bias within one survey may be neutralised by random allocation of subjects to observers. Each observer should also be identified by a code number on the survey record; analysis of results by observer will then identify major problems, and perhaps permit some statistical correction for the bias.

Random subject variation—When measured repeatedly in the same subject, physiological variables like blood pressure tend to show an approximately normal distribution around the subject's mean. Nevertheless, surveys usually have to make do with a single measurement, and the error will not be noticed unless the extent of subject variation has been studied.

Random subject variation has some important implications for screening and also in clinical practice, when subjects with extreme initial values are recalled. Thanks to a statistical quirk this group then appears to improve, for its members include some whose mean value is normal but who by chance were above this value at first examination: on average, their follow-up values necessarily tend to fall (*regression to the mean*). The size of this effect depends on the amount of random subject variation. Misinterpretation can be avoided by repeated examinations to establish an adequate baseline, or else (in an intervention study) by including a control group.

Biased (systematic) subject variation—Blood pressure is much influenced by the temperature of the examination room, as well as by less readily standardised emotional factors. Afternoon surveys of patients with diabetes find a much higher prevalence than morning surveys; and possibly the standard bronchitis questionnaire elicits more positive responses in winter than in summer. Thus conditions and timing of an investigation may have a major effect on an individual's true state and on his responses. Such sources of biases are often unrecognised, and even when recognised they may be hard to control.

Implications of variability

The practical consequences of variability depend less on whether the error is due to observer or subject than on whether it is random or biased.

RANDOM ERROR

(1) Individuals are apt to be misassessed or misclassified. This is serious in clinical practice and in screening, both of which are concerned with the correct management of individuals; but it is not in itself serious in epidemiology, which is concerned with correct decisions about groups.

(2) The mean value for a group is estimated with less cer-

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

tainty (that is, its standard error is increased), but fortunately this uncertainty may be reduced to *any required extent* simply by increasing the sample size, or by replicating measurements in individuals. The strength of epidemiology (and one that clinicians may be inclined to question) is its ability to tolerate crude and inaccurate measuring techniques, provided that sample size and study design are adequate.

(3) An important exception to epidemiology's ability to tolerate random individual errors is in the search for within-subject associations, such as genetic and other risk factors. Random error in the measurements will necessarily tend to cloud true associations and to underestimate true correlations: if crude methods can demonstrate a significant correlation, one may be reasonably sure that its real magnitude is greater than appears.

(4) The size of all these errors from random variation can be estimated, provided that the investigator included an inquiry into measurement variability.

All error in measurements is undesirable; but, provided its

size is known, random error can generally be tolerated. It is "clean dirt."

BIASED ERROR

(1) By definition this biases conclusions about groups (means or rates) and distorts comparisons. Epidemiologically, this is fatal.

(2) Its effects are in no way reduced by an increase in sample size, a large study that is biased being no better than a small one. (Indeed, it is worse, because it is more likely to be believed.)

(3) It is often unrecognised or hard to quantify. Efforts to compensate for it in the analysis must be viewed with suspicion.

Biased error strikes at the roots of comparisons. It is "dirty dirt," and in epidemiology it cannot be tolerated.

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

MATERIA NON MEDICA

Green Mountains

There was my man just ahead, sitting with his back against a big boulder. "Hello!" I shouted. He did not move. Must be asleep! I came round to the front of the rock. There was no need of a second look. The Englishman was dead. There was a burnt out cigarette stub between his fingers. His right shoe and sock were off. He had been bathing a smashed ankle in a pool beside him; his face had injuries. He was facing—gazing it seemed—down the gorge to where lay safety and civilisation, towards which he had unerringly gone from the moment he had left the wreck.

These words come from a book called *Green Mountains* written by Bernard O'Reilly, a member of a pioneering family brought up in dense bush in south-east Queensland. In this book he describes the tragic events of February 1937, when a Stinson airliner crashed in the MacPherson Ranges, a victim of one of Queensland's greatest enemies—the cyclone. It had been blown completely off course to crash in the dense subtropical rainforest of the Lamington Plateau, which lies inland from what is now Australia's major tourist centre on the eastern seaboard—Surfer's Paradise. The storm had been so bad that all contact had been lost. No one knew where the missing airliner had come to ground.

For the week that followed there was enacted one of the most intensive aerial searches in Australian history. The effort proved fruitless and it was decided to abandon the search. Nine days after the crash, inspired by a hunch—if such a thing exists—Bernard O'Reilly set out alone to the area where he thought it most likely the Stinson had landed. Due to his tenacity the crash was located and the lives of the two survivors were saved.

This story of one man's brave efforts has never been forgotten, and to this day bush walkers walk through the rainforest edging Christmas Creek. They go first of all to find the Englishman's grave, which is situated in a grove made shady by tall trees and floored by ferns and orchids. Only a brave few tackle the difficult ascent to the air crash site, the plateau being over 3000 feet and access up a steep, slippery path through rocks and dense undergrowth.

Today, a few pieces of tangled, rusted metal and a small cross are all that remain to mark a tragedy and a supreme challenge by one man. —J P COLQUHOUN (Cleveland, Queensland, Australia).

Job satisfaction

Package tours are socially acceptable these days, even coach ones. The latter range from the sociable and rowdy to the discreetly luxurious, with all mod cons and fares to match. Our package included air flight to Los Angeles and then a two thousand mile round trip on the west coast. The cost was considerable and the passengers largely English, mostly over middle age, relaxed, and able to indulge themselves with this sort of holiday. There were two much younger men travelling together, one a truck driver and the other an engine driver,

the latter intelligent, articulate, and delighted to talk about his job. When he was a boy living in the Midlands the engine sheds were at the bottom of his road and his one ambition was to be a driver. Accepted as a stoker when older, he rose to firing the great Pacific locos. With their gradual phasing out, he was selected for a driver's course and now drives one of the crack intercity trains. The expertise acquired tends to be taken for granted, at least by passengers, but is interesting to hear about, by those curious about other people's jobs: for example, the need to sense the weight of the train in terms of passengers, which calls for a nice judgement in deceleration and braking to bring it neatly alongside the length of the station—apparently one of the trainee driver's passing out tests.

Leaving the Yosemite National Park our coach stopped at a railway museum with an engine of highly specialised type that had been used on the defunct Hetch Hetchy railway, which once served the area. In a trice our companion was crawling monkey-like over it, fingering the controls and reading the name plates—a happy man clearly, and not one feeling the bitterness of an aim achieved. I can understand it well enough when drifting along the Thames above Marlow to see the West Country fliers pounding over Brunel's low brick bridges with coaches cream and blue. Noting my excitement my wife tells me I have never grown up, but alas there is little time left for that now.

Whether airline pilots get the same thrill is debatable; for them seniority and experience is measured by the number of times they "get off the ground"—the pilots' own term—with tedium thereafter as the autopilot takes over. Yet I suppose there must be thrills too; in the last hour of the Great Circle route to Los Angeles the technicolour sunset above the clouds may be reward enough, then the descent into the inky blackness beneath, with the bright lights of Las Vegas to guide the final run in. Even so, I would prefer to be the engine driver with Reading safely behind, and signals green all the way.—ROBERT CUTLER (retired consultant dental surgeon, Long Ditton, Surrey).

How soon after a suspected attack of glandular fever can a young adult be immunised for tropical travel?

No inoculation should be carried out unless a patient has fully recovered from an infection such as glandular fever but as your patient now feels well, and the blood picture is normal, the inoculation schedule can proceed. A yellow fever vaccination should be performed at one of the designated vaccination centres as soon as possible, and after this a booster dose of TAB/tetanus vaccine, either 0.5 ml subcutaneously or 0.1 ml intradermally, cholera vaccine in a dose of 0.5 ml, and the oral poliomyelitis booster could all be given together. In a further two to four weeks the second cholera dose of 1 ml subcutaneously should be given.