Epidemiology for the Uninitiated

Planning a survey

GEOFFREY ROSE, D J P BARKER

British Medical Journal, 1978, 2, 1142-1143

Medical planning requires information on the extent and distribution of health problems; often this is not available, especially locally. A geriatrician asks about the extent of unmet need among district residents. An inner-city psychiatrist wants to back his request for more facilities by some evidence of high local morbidity. A physician needs to know how much hypertension would be uncovered if blood pressures were measured in non-medical clinics. A general practitioner wonders if there really is a local excess of bronchitis. A chemical pathologist wishes to establish a "range of normal." In some such ways almost any doctor may find himself needing to undertake a simple cross-sectional survey. We will attempt to help him in our articles this week and next.

Early planning

Actual data collection begins only late; its success depends on careful preparation, often requiring many months.

BACKGROUND READING

Library research may uncover unsuspected sources of published information (the Registrar General's mortality and cancer registry reports, Hospital Inpatient Inquiry, Morbidity in General Practice reports); special tabulations may be provided on request to the Hospital Activity Analysis; similar surveys may have been done elsewhere. Even if the library does not yield the whole answer it often guides the planning.

DEFINING SPECIFIC QUESTIONS

The first and often the most difficult question is "Why am I doing this survey?" Many studies start with a general hope that something interesting is sure to emerge, and these ones often end in frustration. The general interest has first to be translated into precisely formulated, written objectives. Every survey should be reasonably sure to give an adequate answer to at least one specific question. This initial planning requires

some idea of the final analysis; and it may be useful at the outset to outline the key tables for the final report, and to consider the numbers of cases expected in their major cells.

Every study needs a primary purpose. It is easy to argue, "While we have the subjects there, let's also measure . . ."; but overloading, whether of investigators or subjects, must be avoided if it in any way threatens the primary purpose. Sometimes subsidiary objectives may be pursued in subsamples (every *n*th subject, or in a particular age group), or by recalling some subjects for a second examination: when their initial contact has been favourable, response to recall is usually good.

Choice of examination methods

The overriding need in an epidemiological survey is to examine a representative sample of adequate size in a standardised and sufficiently valid way. This determines the choice of examination methods and the points where these differ from those of clinical practice. Methods must be acceptable, and if possible non-invasive, or co-operation suffers and the study group becomes unrepresentative. They must be relatively cheap and quick, or not enough subjects can be examined: with fixed resources the need for detail conflicts with the need for numbers. Most important of all, methods and observers must be capable of rigorous standardisation, even if this excludes the benefits of clinical judgment.

The price of these necessary characteristics of epidemiological methods may be a substantial loss of validity (sensitivity or specificity or both) and, in the interests of making fair comparisons, conclusions about individuals in the study will be limited.

Sampling

SAMPLE SIZE

Most surveys and trials are smaller than the investigator would wish, lack of numbers often setting a limit to some desirable subgroup analysis. This is inevitable. What can be avoided is discovering only at the final analysis that numbers do not permit achievement even of the study's primary objective. To prevent this disappointment the purpose of the study has first to be formulated in precise statistical terms. If the aim is to estimate prevalence, then sample size will depend on the required accuracy of that estimate. (The table gives some examples.) You will notice that sampling error is proportionally greater for less common conditions; that is to say, to achieve the same level of confidence requires a larger sample if prevalence is low. Doubling the size of the study does not double its accuracy; in fact, it reduces the standard error by a factor of $\sqrt{2}$.

Techniques also exist for calculating sample sizes required for estimating, with specified precision, the mean value of a variable, or identifying a given difference in prevalence or mean

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

95 % confidence limits for various rates and sample sizes

Estimated prevalence (%)	95 % confidence limits	
	n = 500	n = 1000
2	1.0-3.7	1.2-3.1
10	7.5-13.0	8.2-12.0
20	16.6-23.8	17.6-22.6

values between two populations. These techniques may be found in the textbooks or (better) by consulting a statistician; but either way the investigator must first know exactly what he wants to achieve.

SAMPLING METHODS

Statistical inference is possible only if the sample is random, or effectively random; that is to say, if each individual in the study population has a known (usually identical) probability of selection. To achieve this a *census* or listing of the study population is first required. In a survey of adults in a hospital district the electoral register will probably serve. In an occupational group the payroll is invariably complete, and in a school there are class registers. In general practice there may be an age-sex register (although many are inaccurate); otherwise a *systematic*

sample may be obtained by taking the notes of every *n*th patient, and this should be adequately representative.

In a *simple random sample* the listed subjects are numbered serially. Numbers within the appropriate range are then read off from a table of random numbers until enough subjects have been chosen.

It may be that the investigator wishes to choose a sample in which certain subgroups (particular ages, for instance, or high-risk categories) are relatively overrepresented. To achieve this he may divide the study population into subgroups (*strata*) and then draw a separate random sample from each, but adjust the various sample sizes to suit his requirements. This is a *stratified random sample*.

The study population may be large and widely scattered—for example, all the general practices in a city—but for the sake of convenience the investigator may wish to concentrate his survey in a few areas only. He can do this by drawing first a random sample of practices, and then, within these practices, drawing a random sample of individuals. This is *two-stage sampling*. There is some loss in statistical efficiency, especially if at the first stage only a few units are selected. If the acceptable number of first-stage units is very small then random selection is unsuitable, judgment serving better than chance.

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

General Practice Observed

Early treatment of myocardial infarction in the community

F G DUNN, S ELMS, D I MELVILLE, T D V LAWRIE, T S MURRAY

British Medical Journal, 1978, 2, 1143-1145

Summary and conclusions

The prehospital management of acute myocardial infarction by general practitioners and emergency-treatment service physicians was analysed in 53 patients. The correct clinical diagnosis was made or suggested before admission in 47 patients. Only 25 patients received analgesics from the general practitioner, and 32 were still in pain and needed diamorphine on admission to hospital. Only one patient received antiarrhythmic treatment.

The findings suggest that, despite accurate clinical diagnosis of acute myocardial infarction, deficiencies

exist in the prehospital phase of management and that education programmes are worth consideration.

Introduction

Early studies in the evaluation of coronary care units showed that mortality from acute myocardial infarction had improved when compared with general medical wards.¹⁻⁴ Recently, however, the exact value of coronary care units has been seriously questioned, since it is well known that most patients who die from acute myocardial infarction do so before reaching hospital.⁵ The coronary care ambulance is an important development in bringing early coronary care to patients,^{6,7} but there are both practical and economic problems in instituting this system on a widespread basis.⁸ Thus the general practitioner should take a more active role in the critical early phase of management. We aimed at assessing to what extent this was already being done and at analysing patterns of management of acute myocardial infarction by general practitioners.

Patients and methods

Fifty-three patients (40 men and 13 women) who were admitted to the coronary care unit of the Glasgow Royal Infirmary were studied. Two criteria were used for selection: (1) all patients had been seen by either their general practitioner (40 patients; 76%) or an

University Department of Medical Cardiology, Royal Infirmary, Glasgow G4 0SF

F G DUNN, MB, MRCP, senior registrar

S ELMS, MB, CHB, senior house officer

D I MELVILLE, MB, MRCP, registrar

T D V LAWRIE, MD, FRCPED&GLAS, Walton professor of medical cardiology

University Department of General Practice, Royal Infirmary, Glasgow G4 0SF

T S MURRAY, MRCP, senior lecturer