

Statistical Notes VI: Study Design

A Sheikh, A Cook

INTRODUCTION

Thus far in this series we have concentrated on issues to do with summarising data, estimation (inference) and testing associations. A competent statistical analysis will not, however, rescue a study that is either poorly designed or executed. In this paper, we consider some of the issues pertinent to the design of studies, focussing on the need for clear study objectives, choosing suitable outcome measures, and deciding on an appropriate study methodology to answer the question(s) being posed. The choice of methodology will influence the size of the study, the data items that need to be collected, and also the plan of analysis.

CLEAR STUDY OBJECTIVES

Arguably the most important aspect of designing a study is being clear about the question(s) which it is hoped the study will answer. Questions most commonly posed can be classified under three broad headings:

- **Estimating** certain population characteristics, such as the prevalence of asthma in a geographical health authority at a particular point in time (point prevalence)
- Identifying **associations** between exposures and outcomes, such as the relationship between inhaled corticosteroid use in children and final height
- Evaluating the **efficacy** and/or **effectiveness** of an intervention, such as pneumococcal vaccination in reducing incidence of pneumococcal pneumonia in adults with asthma.

These three questions sometimes follow each other in order: the scale of a problem needs to be assessed first, contributory factors are then identified, and finally potential solutions can be tested.

CLEAR OUTCOME MEASURES

Another key consideration is the choice of suitable outcome measures. Of particular importance is the need to distinguish between so-called 'primary' and 'secondary' outcome measures. Primary outcome measures are used as the basis for sample size calculations and should, ideally, satisfy certain characteristics. They should be clinically relevant, reliably measurable, and comparable with other studies (increasingly important with the development of meta-analytic techniques). In a study of pneumococcal vaccination efficacy for example, a suitable primary outcome measure might be episodes of microbiologically or serologically confirmed pneumococcal pneumonia in the one-year period following vaccination.

It can sometimes be tempting to measure surrogate outcomes; this temptation should however be resisted,

unless the surrogate measure is known to be a valid and reliable predictor of the outcome measure of interest. In a study designed to determine the efficacy or effectiveness of pneumococcal vaccine in preventing pneumonia for example, a surrogate outcome that is sometimes used is serological evidence of immunity following vaccination. The difficulty with such measures however lies in interpreting the clinical relevance of these findings. With very rare outcomes (for example asthma death) there may however be no alternative to the use of surrogate outcomes.

CHOICE OF STUDY DESIGN

The choice of study methodology should be determined on the basis of considerations such as the question being asked, available resources (financial and human), and designs that reduce the risk of systematic error (bias and confounding, see table 1). Study designs commonly used in health services research can be classified under two broad headings: observational and experimental studies.

Observational studies

Observational studies are essentially of three types:

- Cross-sectional studies
- Case control studies
- Cohort studies

Cross-sectional studies provide a 'snapshot' picture at one point in time. They are therefore useful for quantifying the scale of a problem, such as the prevalence of smoking in hospitalised chronic obstructive pulmonary disease (COPD) patients on a particular day of the year. Cross-sectional studies have the advantages of being relatively quick, cheap and straightforward to analyse providing information on associations between exposure and disease (smoking and COPD for example). No information is however provided on causation and it is here that cohort and case-control study designs are particularly useful.

Cohort studies follow a group of people through time to determine the proportion who develop the outcome of interest. Classically these will be disease free individuals, some of whom are 'exposed' and others not 'exposed' to the phenomenon of interest and the proportion who are affected in each group are determined (for example the development of lung cancer in smokers and non smokers). They are sometimes referred to as prospective studies since people are identified in advance and followed up 'prospectively'. This term is however best avoided since it is also possible to conduct retrospective cohort studies using records that predate the onset of a condition and following these through time to compare disease occurrence between exposed and unexposed people. From cohort studies it is possible to obtain information on a temporal relationship

Aziz Sheikh
NHS R&D National
Primary Care Training
Fellow

Adrian Cook
Statistician

Correspondence to:

Aziz Sheikh
Department of Primary
Health Care & General
Practice
Imperial College School
of Medicine
Charing Cross Campus
St Dunstan's Road
London W6 8RP
aziz.sheikh@ic.ac.uk

Date Submitted: 01:06:01
Date Accepted: 11:06:01

Prim Care Respir J 2001;
10(2);50-51

Table 1: Sources of Systematic Error

- Selection bias refers to any error in selecting the study population such that the people who are selected to participate are not representative of the reference population or when the groups under study are not comparable.
- Information bias refers to any error in the reporting or measurement of exposure or outcome that results in systematic differences in the accuracy of information collected between comparison groups.
- Confounding occurs when an estimate of the association between an exposure and an outcome is also affected by another exposure on the same disease, and the two exposures are correlated.

between exposure and disease and also obtain an estimate of the incidence of a condition between groups thus making it possible for a dose-response relationship to be established. The major disadvantages are that cohort studies can be slow and expensive, particularly for rare diseases, which require a large amount of observational time. Analysis of cohort studies is also often complex.

Case control studies are better for studying rare diseases or uncommon events (such as asthma deaths) but the comparator group needs careful selection to avoid introducing selection bias. In these studies, people are classified on the basis of disease status. Attempts are then made to obtain an estimate of exposure to the factor(s) of interest between the two groups. Recall (information) bias is an important concern, as is the difficulty of controlling for confounding factors. Despite these reservations, if well conducted, case control studies offer several advantages including quick results, efficiency, and relative ease of analysis.

Experimental studies

The distinguishing feature of experimental studies is that the investigator assigns subjects to the different groups being compared; herein lies the main advantage because the risk of confounding (see table 1) is greatly diminished. Randomisation confers the additional benefit of controlling for all (known and unknown) confounding factors and it is for this reason that randomised trials represent the methodology of choice for evaluating the efficacy and effectiveness of interventions. A number of trial designs now exist including parallel group, cross-over, factorial, and cluster ran-

domised trials; discussion of the relative strengths and weaknesses of each trial design fall beyond the scope of this paper and interested readers are referred to the detailed critique provided by Pocock (further reading). Irrespective of the particular trial design chosen, subjects and assessors should be blinded to assigned treatment wherever possible since this will in addition minimise the risk of information bias.

CONCLUSIONS

In this paper, we have highlighted the close inter-relationship between issues to do with the design of studies and their subsequent analysis and interpretation. Our experience suggests that incorporating a detailed plan of statistical analysis into the study protocol stage will often help in clarifying the key issues that need to be considered. ■

ACKNOWLEDGEMENTS

Aziz Sheikh is supported by a NHS R&D National Primary Care Award. We thank Dr Robert McKinley for his constructive suggestions on an earlier draft of this paper.

FURTHER READING

1. Coggon R, Rose G, Barker DJP. *Epidemiology for the uninitiated*. London: BMJ Books, 1997:27-51.
2. Hennekens CH, Burling JE. *Epidemiology in medicine*. Boston: Little Brown, 1987:101-213.
3. Kirkwood BR. *Essentials of medical statistics*. Oxford: Blackwell Science, 1998:154-90.
4. Pocock SJ. *Clinical trials. A practical approach*. Chichester: Wiley, 1983.