Bias in cohort studies

Huw TO Davies, Iain K Crombie

Cohort studies allow an exploration of patient change over time. They can provide information on the incidence of disease, prognosis (including patient satisfaction) and likely health-care resource use. Nonetheless, bias can be present in cohort studies in the way patients are selected and followed-up, the way measures are taken, or the way data are analysed. This short paper explores ways in which such flaws can be uncovered in published studies, so that their findings can be interpreted appropriately.

INTRODUCTION

Many worthwhile questions in health care involve an exploration of how events unfold over time. What is the incidence of disease? What prognosis can patients expect, and how does this vary depending upon the patients' characteristics and circumstances? What health-care services will patients use at different points in their illness? What binds these questions is that they all involve an unfolding picture; attempts to provide answers to these questions therefore involve summarizing events over a period of time.

The research design most appropriate to answer these sorts of question is called a cohort study (also known as a follow-up study or a longitudinal design). In essence, a cohort study involves selection of a group of individuals, followed by measurement of some aspects of those individuals on at least two occasions separated by a time interval. Sometimes repeated measures are taken at different intervals widely spaced — in the order of months, years or even decades. In other studies just two measures are taken and the time elapsed might be as short as days or even hours.

Dr Huw TO Davies is Reader in Health Care Policy and Management, Department of Management, University of St Andrews, St Andrews, Fife KY16 9AL and **Professor Iain K Crombie** is Professor of Epidemiology and Public Health, Department of Epidemiology and Public Health, University of Dundee, Ninewells Hospital and Medical School, Dundee

Correspondence to: Dr HTO Davies

The word cohort is derived from the name of a unit of soldiers in a Roman legion: a cohort being a tenth part of a legion. As a cohort of soldiers marches along a road, so study subjects in a cohort study can be thought of as passing through time, from one health state to another. Data collection in a cohort study is then analogous to capturing information on individual soldiers as they pass specific way-markers.

Despite the basic simplicity of design, bias can creep into cohort studies in a number of ways. This short article explores the source of these biases and advises on their identification and assessment.

THE RIGHT DESIGN

Cohort studies are used to explore change, and in answering questions like this they are far superior to surveys (Davies, 1999a). However, analyses reporting changes in individuals are often embedded in survey projects. So the first question to ask of any research reporting relationships over time is: was the design a true cohort with repeated measures on the same sample group collected at different time intervals? If data were in fact collected at only one time point, with patients asked to recollect facts from earlier periods, then the design is a survey not a cohort — and surveys are not well suited to this kind of analysis (Davies, 1999a).

Some research studies report data collected on *different* groups of individuals collected at distinct time points, and try to draw conclusions about change. Such a design is again not a cohort study (which involves repeated data collection on the *same* group of individuals). It is instead simply a collection of surveys. Again, the interpretation of data from repeated surveys is problematic.

Even when the study design is a true cohort, it may be poorly suited to answering the research question posed. In particular, assessing whether changes in patient status are causally related to health-care activities needs the added design features of a rigorous trial (Davies, 1999b) if the dangers of confounding are to be avoided (Davies and Williams, 1999). For example, cohort studies of the effects of maternal analgesia during childbirth on subsequent behaviour of the baby have found many interesting associations (Lieberman et al, 1979). However, interpretation of these is difficult because of the many maternal factors which influence both choice of maternal analgesia and subsequent baby development. Cohort studies can *describe* changes and the factors that are associated with these changes; they are less able to *explain* whether these factors are causal or not (Davies and Williams, 1999).

SAMPLE SELECTION

Cohort studies begin by taking a sample of patients. Thus many of the issues about representativeness discussed in a previous article on surveys (Davies, 1999a) are equally germane to the sampling of individuals in cohort studies. Selection bias takes many forms (Sackett, 1979) and can lead to unusual groups and artefactual relationships. For example, samples of patients recruited from specialist clinics will show important differences from those recruited from community settings (Crombie and Davies, 1998). They will tend to have more severe and intractable disease and will have had longer treatment histories. These differences in turn may have a marked influence on the findings reported.

MEASUREMENT

Once the sample has been selected, issues arise over measurement. Are the measures used valid and reliable? Were all the important and relevant characteristics of the individuals measured during the initial data capture? Were all the appropriate outcomes captured during the second phase of data collection?

Omissions and failures here can seriously undermine any cohort study. For example, a study that investigated the prognosis for patients with lung disease would need to collect baseline data on smoking habits and environmental exposure to atmospheric irritants. Any failure to do so would leave the findings seriously deficient in terms of their interpretability.

COMPARISON BETWEEN GROUPS

Cohort studies are frequently concerned with making comparisons between different subgroups defined by some characteristic at the initial time point. For example, we might be asking whether smokers fare better or worse than non-smokers in their disease progression, or whether men differ from women in their use of health-care resources, or whether those treated on an inpatient basis do better than those seen as outpatients.

While such descriptive analyses may be informative, moving from using variables to describe a relationship to using them to explain that relationship can be a slippery slope. As described in a previous paper in this series, associations between variables do not necessarily indicate that the relationship is causal (Davies and Williams, 1999). The presence of confounding variables can both generate spurious connections and hide real relationships.

A further difficulty in making comparisons occurs when many different subgroups are compared in the search for significant differences. Unless comparisons are relatively few in number and have been pre-specified, such fishing expeditions violate the basic assumptions underlying statistical hypothesis testing (Davies, 1998a).

FOLLOW-UP

The inclusion of a second phase of data collection on the original group of study subjects differentiates cohort studies from surveys. As this additional data gathering may take place many months, years or even decades after the first, there is great potential for individuals to get ‘lost’ to this follow-up. Such losses to follow-up pose important challenges to the validity of cohort study findings.

The problem is that those who fall by the wayside may be different in important and systematic ways from those who remain in the study. For example, a study that followed up patients after treatment to assess their satisfaction with the services provided might fail to collect data on those who do not keep their clinic appointments. Such losses to follow-up clearly have the potential to bias study findings.

A second problem with follow-up is that sufficient time should have elapsed to allow any phenomena of interest to emerge. Follow-up that is too short may fail to uncover important relationships. Thus a study looking at patient recovery from surgery may miss important chronic discomfort if follow-up is limited to just a few weeks (Macrae and Davies, 1999). The problem with long-term follow-up is that it provides additional opportunities for significant numbers of study subjects to be lost.

ANALYSIS PITFALLS

The key analysis of cohort data involves describing patient outcomes at the second time point and relating these to the patients’ initial characteristics. There are at least three important ways in which such an analysis can be incomplete or misleading. First, the analysis should take into account any confounding variables and adjust for these. For example, a cohort study that found an apparent relationship between coffee drinking and heart disease would need to adjust the findings for differences between coffee drinkers and non-coffee drinkers in terms of smoking habits, dietary intake and a host of other lifestyle factors. It can be quite difficult to know whether all potential confounders have been adjusted for and this area of the discussion in any published report needs special scrutiny.

A second pitfall in the analysis of cohorts lies in the potential for contemporaneous changes to cloud the findings. One specific instance of this is that patients age as time passes, and therefore any analysis must take account of this. This is only likely to be a problem when the period of follow-up is significantly long. For example, a study that examined the long-term incidence of disease for different occupational groups would need to take into account the fact that many diseases increase with age, and that many occupational groups have different age profiles.

Third, many cohort studies recruit study subjects over a period of time. This means that there may be different lengths of follow-up on different individuals, which must also be accounted for in the analysis.

Finally, any analysis of cohort data needs to take account of the play of chance (Davies, 1998a). All the measures used to describe the initial cohort can (and should) be presented in terms of both their average and their spread. Sometimes this will use means and standard deviations, at other times medians and percentiles will be more appropriate (Davies 1998c). In addition, the main measures used in cohort analyses (usually relative risks — Davies, 1998b) can all be calculated with confidence intervals. This approach has an important advantage over traditional hypothesis testing in that it presents a range of possible scenarios that are compatible with the empirical data rather than simply denoting statistical significance (Gardner and Altman, 1986).

RETROSPECTIVE COHORTS

We have described cohort studies in terms of the selection and measurement of patients, the elapse of time, and then further data collection on the same group of individuals. Such studies, when carried out in this manner, are called prospective cohorts.

There is, however, another way in which cohort studies may be carried out which is conceptually identical but practically advantageous. It concerns an approach to minimize the impact of protracted follow-up. All that waiting around for time to elapse during a cohort study is enough to try anyone's patience — and so an elaboration on the design has evolved. Called retrospective cohorts, these studies reach back into the past to define a cohort at some historical time point. The initial data on this cohort are then collected from historical records of some sort, and subsequent data are gathered either from existing records or through new contacts being made with the individuals. For example, in designing a study relating birth weight to subsequent intellectual development, we might define a cohort as: 'all babies born in a random sample of UK hospitals in 1971'.

Initial data collection would then be captured from old medical records. For the second phase of data collection, sampled individuals could either be traced and interviewed, or school records could be searched to find educational attainment.

The obvious advantage of retrospective cohort studies is the obviation of the need to wait around for time to pass between data collections. Equally obviously, the approach is heavily dependent on the availability, accuracy and completeness of historical data sources. Thus the key pitfalls in retrospective cohort studies are the same as those for prospective studies only writ large: selection bias, measurement deficiencies and losses to follow-up.

CONCLUSIONS

Cohort studies foster an understanding of change in patients, and the relationship of that change with patient and service characteristics. The opportunity to avoid bias and erroneous interpretations makes cohort studies superior to surveys in answering such questions but inferior to randomized controlled trials in assessing cause and effect. Nonetheless, there are many instances when randomized controlled trials are impractical, unethical or otherwise undesirable and thus observational studies like cohort designs have an important role to play (Black, 1996).

Many of the potential biases in survey design are also relevant considerations for cohort studies (Davies, 1999a). In addition, further pitfalls lie in inadequate follow-up (in terms of time or completeness), and erroneous analyses (especially lack of attention to confounding factors, including contemporaneous change). More detailed explanations of the design and analysis of cohort studies, and their critical appraisal, can be found in a range of useful texts (Sackett et al, 1991, 1997; Laupacis et al, 1994; Crombie, 1996). HM

KEY POINTS

- Cohort studies answer questions about what happens to patients over time, and how this experience differs between various patient groups.
- Cohort studies involve two (or more) distinct phases of data collection on the same individuals separated by the passage of time.
- Although data collected in surveys (at a single time point) may be analysed in a longitudinal manner, such analyses are weak and potentially flawed.
- The key areas where bias can creep into cohort studies are: when the patients are selected, when the measures are taken, when patients get lost to follow-up and when analyses do not properly consider confounding.
- Retrospective cohort studies may be efficient ways of gathering data for long-term studies but they are especially prone to all the main pitfalls found in any cohort design.

Black N (1996) Why we need observational studies to evaluate the effectiveness of health care. *Br Med J* **312:** 1215–8

Crombie IK (1996) *The Pocket Guide to Critical Appraisal.* BMJ Publishing, London

Crombie IK, Davies HTO (1998) Selection bias in pain research. *Pain* **74:** 1–3

Davies HTO (1998a) Assessing chance variability in treatment trials. *Hosp Med* **59:** 650–2

Davies HTO (1998b) Interpreting measures of treatment effect. *Hosp Med* **59:** 499–501

Davies HTO (1998c) Informative presentation of summary data. *Hosp Med* **59:** 154–5

Davies HTO (1999a) Bias in surveys. *Hosp Med* **60:** 898–900

Davies HTO (1999b) Bias in treatment trials. *Hosp Med* **60:** 599–601

Davies HTO, Williams FLR (1999) Confounded by confounding: separating association from causation. *Hosp Med* **60:** 294–7

Gardner MJ, Altman DG (1986) Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br Med J* **292:** 746–50

Laupacis A, Wells G, Richardson WS, Tugwell P (1994) Users' guides to the medical literature. V. How to use an article about prognosis. *JAMA* **272(3):** 234–7

Lieberman BA, Rosenblatt DB, Belsey E et al (1979) The effects of maternally administered pethidine or epidural bupivacaine on the fetus and newborn. *Br J Obstet Gynaecol* **86:** 598–606

Macrae WA, Davies HTO (1999) The epidemiology of chronic post-surgical pain. In: Crombie IK, Linton S, Croft P, Von Korff M, LeResche L, eds. *The Epidemiology of Chronic Pain.* International Association for the Study of Pain, Seattle: 125–42

Sackett DL (1979) Bias in analytical research. *J Chronic Dis* **32:** 51–63

Sackett DL, Haynes RB, Guyatt GH, Tugwell P (1991) *Clinical Epidemiology: A Basic Science for Clinical Medicine.* Little, Brown and Company, Boston, Massachusetts

Sackett DL, Richardson WS, Rosenberg W, Haynes RB (1997) *Evidence Based Medicine: How to Practice and Teach EBM.* Churchill Livingstone, London