Bias in case-control studies

Huw TO Davies, Iain K Crombie

Case-control studies are largely used to explore differences between groups of individuals. They can identify potential risk factors associated with disease, or they can investigate patient behaviour, such as why some people do not attend for services. As such, case-control studies are often used to generate or test hypotheses about causal factors. Nonetheless, bias is always a danger in case-control studies, arising especially from the way in which study samples are selected or from the collection of retrospective data. Confounding also remains a problem. This short paper explores ways in which such flaws can be uncovered in published studies.

INTRODUCTION

Epidemiology is concerned with identifying and assessing causal factors related to disease; clinical epidemiology in turn addresses the amelioration of that disease as a result of interventions. Both branches of study use a variety of research designs to uncover these causal relationships between patient characteristics, experiences and health service interventions on the one hand, and health outcomes on the other.

Randomized controlled trials offer the strongest evidence of cause and effect (Davies, 1999), but in many circumstances they can be unethical or impractical. Cohort studies, as discussed previously (Davies and Crombie, 2000), can also help unravel what happens to patients and why, but again problems in execution and interpretation abound (Davies and Williams, 1999).

There are many reasons why prospective studies such as trials or cohort studies might be impractical, inappropriate or unethical. For example, assessing the damaging effects of exposure to various pathogens clearly cannot be countenanced as an experimental design — one cannot imagine randomly allocating patients to smoke 20 cigarettes a day. Even passive observation of potentially harmful exposures poses serious ethical dilemmas. In addition, long-term or very rare outcomes may be impractical to study using prospective designs. In such situations case-control studies offer some advantages. They are relatively cheap and, because they focus on events that have already occurred, relatively quick to conduct. In addition, they can assess the importance of a wide range of possible causative factors in a single study.

The essence of case-control studies is simple. A group of 'cases' are selected — usually individuals with a specific disease — and these cases are then compared to a group who do not have the disease or other defining characteristic of the cases (these are the controls). Differences between the two groups are then explored in an attempt to 'explain' how cases get to be cases. Thus case-control studies are retrospective — looking backwards from a condition to its antecedents. This is in fundamental contrast to trials and cohort studies, which move forwards from exposure to outcomes.

Most often, case-control studies are used to explore disease aetiology or the harmful impacts of various exposures or lifestyles. More rarely they have been used to examine the long-term benefits of interventions such as screening or vaccination. In principle case-control studies can be used to explore what it is that makes any group different from another — whether the groups are defined by disease status or by some other marker.

For all the basic simplicity of design, case-control studies are rather prone to bias — perhaps even more so than other observational studies such as cohort designs. Bias can arise in three different areas:

- In the ways in which both cases and controls are selected
- In the ways in which measures are taken
- In the potential for confounding.

This short article explores these biases and advises on their identification and assessment.

THE RIGHT DESIGN

The first issue when confronted with a published report of a case-control study is to ask whether the approach used is the most appropriate. For example, would a prospective cohort study be practical and offer better information? Often the answer to this question is 'no' — especially when the outcomes examined are rare, long-term or unexpected. In such cases, a case-control study may be the best that can be expected. However, if the issue under examination can be studied by other means, then the findings from case-control studies should be regarded as preliminary — more hypothesis raising than definitive findings.

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

SAMPLE SELECTION

The individuals included in a case-control study consist of two separate groups: the cases and the controls. Selection of each poses distinct problems. First there is the problem of defining 'caseness': what is it that makes a case a case? Often, when examining disease aetiology, a case is defined by some diagnostic criteria. In other studies different characteristics will be used, for example, 'non-attenders' in studies examining why people miss appointments, or 'employment status' in studies examining the factors associated with career success. The crucial issue is that the definitions should be clear-cut and consistently applied: random misclassification errors in the cases can dilute possible findings, whereas systematic misclassification errors can lead to specious results.

The second problem in selecting cases is the need for these to be in some senses 'typical'. If the cases selected are highly atypical then any associations discovered may not generalize — limiting the usefulness of the research. For example, cases selected from tertiary referral centres may differ in important ways from those with the same disease seen in general practice.

Cases are only one half of a case-control study, and it is in the selection of controls that the biggest potential pitfalls lie. For comparisons to have any validity, the controls need to be as similar as possible to the cases — except, of course, that controls should lack the defining features of 'caseness'. Because cases and controls are selected separately from potentially very different populations, the concern is that it is these selection forces which may account for any differences found between the two groups — leading to spurious findings. Thus the crucial question to ask of the controls used is: how might these controls differ in important ways from the cases because of the way in which they were selected?

In order to get a better match between cases and controls it is not unusual for studies to 'match' controls to pre-existing cases on a number of different criteria (e.g. age, sex). While this approach may lead to similar looking groups it does complicate the analysis — which should now take account of this matching process.

MEASUREMENT

Measurement bias poses a serious problem for almost every research study. In case-control studies, the first challenge is to the integrity of the cases themselves. Measurement errors may, for example, cause non-cases to be identified as cases, or vice versa. Thus the first point of scrutiny for measurement errors should be the case definition criteria.

Beyond sample identification, the key role of measurement is to explore possible antecedent factors in both groups. Clearly, to maintain comparability, identical data gathering strategies should be used for both groups. A useful methodological safeguard in this process is 'blinding' of the data gatherers — so that they do not know whether they are collating data for cases or controls. This prevents differential assessments creeping in to the data set.

It is, however, in the retrospective nature of cohort studies that some of the biggest measurement difficulties lie. Study subjects will only rarely be 'blind' to the nature of their condition, and this self-knowledge may impact severely on recollections or assessments of the past. For example, close questioning of parents who have suffered a cot death (the cases) may cause much greater revelation of otherwise minor events (such as mild illness) than would similar questioning of non-bereaved parents (the controls). Thus, whatever the care taken to ensure comparability, recall bias may undermine even well executed case-control studies.

ANALYSIS ISSUES

Case-control studies do not usually draw their cases or their controls from a known population — and therefore cannot provide either incidence or prevalence figures. Thus, unlike cohort studies (Davies and Crombie, 2000), it is not possible to calculate relative risks. Instead, the standard measure of comparison is the odds ratio (Davies, 1998a). This measure is in fact a reasonably good estimate of the relative risk for a wide range of circumstances (Davies et al, 1998).

The unadjusted odds ratio provides only the first look at associations between antecedent factors and caseness. Although controls may have been selected to try to ensure compatibility with the cases there may be many reasons to suspect that they are different. Thus the possibility of confounding remains a real possibility (Davies and Williams, 1999). Adjusting for a wide range of possible confounders produces a new set of adjusted odds ratios which try to take any known confounders into account. Of course, the possibility that there are significant unknown confounders can never entirely be discounted.

Any case-control analysis needs to take account of the play of chance (Brennan and Croft, 1994; Davies, 1998b). The main measures used in the analyses (usually adjusted and unadjusted odds ratios) should all be presented 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). This allows the reader to assess both worst-case and best-case scenarios compatible with the data, thus protecting against over-interpretation. In particular, wide confidence intervals around the odds ratio denote studies which have little power to detect even sizeable real effects.

Finally, one of the strengths of the case-control study — its ability to explore a wide range of possible aetiological factors in a single study — also presents some serious dangers when it comes to statistical analysis. The basic premises of hypothesis testing mean that one comparison in 20 will appear as 'statistically significant' by chance alone. Thus studies which make many comparisons will inevitably 'discover'

some spurious results. This is another reason why the findings of case-control studies should be seen as provisional, unless the study was driven by a limited number of prior hypotheses and the analysis was confined to these factors.

CONCLUSIONS

Case-control studies sometimes suffer from serious biases that may be difficult to avoid or even quantify. The title of a review in this area (*A collection of 56 topics with contradictory results in case-control studies* — Mayes et al, 1988) attests to the potentially misleading nature of the design. Nonetheless, case-control studies do have a role to play in exploring factors associated with rare, unexpected or long-term events. They also provide a valuable service in raising hypotheses that may subsequently be investigated using more robust methods.

The key sources of bias in case-control studies arise from the ways in which samples are selected and measurements are taken. Atypical cases may limit the conclusions that can be drawn from studies, and carelessly drawn controls can vitiate any comparisons. The unavoidably retrospective nature of the design places important limits on the quality of the data that can be gathered and always raises the spectre of recall bias. As with all observational designs the possibility of confounding can never be entirely excluded when interpreting any associations found.

More detailed explanations of the design and analysis of case-control studies can be found in a range of useful texts (Schlesselman, 1982; Sackett et al, 1991, 1997; Levine et al, 1994; Crombie, 1996). Some of these provide handy checklists for critical appraisal of the design. HM

KEY POINTS

- Case-control studies try to answer questions about why individuals with some defining characteristic (often a disease; the 'cases') differ from individuals without that characteristic (the 'controls').
- Case-control studies are unavoidably retrospective in nature moving from some outcome to explore a range of antecedent factors. The quality of evidence they produce is usually considered inferior to that arising from prospective studies.
- Case-control studies may be most appropriate for the investigation of rare, long-term or unexpected outcomes, or for exploring a range of novel aetiological hypotheses.
- Bias in case-control studies arises primarily from the ways in which samples are selected (especially in the choosing of appropriate controls), and from the ways in which data are collected (especially because of the possibility of recall bias).
- Good case-control studies are tentative in their conclusions and feature the following key attributes: clear and reliable definitions of 'caseness'; identical identification and data gathering procedures for controls as for cases; data gathering blinded to the status of cases and controls; an assessment of the potential impact of recall bias; and a comprehensive exploration of the possibility of confounding.

Brennan P, Croft P (1994) Interpreting the results of observational research: chance is not such a fine thing. *Br Med J* **309:** 727–30

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

Davies HTO (1998a) Interpreting measures of treatment effect. *Hosp Med* **59(6):** 499–501

Davies HTO (1998b) Assessing chance variability in treatment trials. *Hosp Med* **59(8):** 650–2

Davies HTO (1999) Bias in treatment trials. *Hosp Med* **60(8):** 599–601

Davies HTO, Crombie IK (2000) Bias in cohort studies. *Hosp Med* **61(2):** 133–5

Davies HTO, Crombie IK, Tavakoli M (1998) When can odds ratios mislead? *Br Med J* **316:** 689–91

Davies HTO, Williams FLR (1999) Confounded by confounding: separating association from causation. *Hosp Med* **60(4):** 294–7

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

Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V (1994) Users' guides to the medical literature. IV. How to use an article about harm. *JAMA* **271(20):** 1615–9

Mayes LC, Horwitz RI, Feinstein AR (1988) A collection of 56 topics with contradictory results in case-control research. *Int J Epidemiol* **17(3):** 680–5

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

Schlesselman JJ (1982) *Case-Control Studies: Design, Conduct, Analysis.* Oxford University Press, New York