Epidemiology in Practice: Case-Control Studies

Susan Lewallen MD
Paul Courtright DrPH
BC Centre for Epidemiology & International Ophthalmology
University of British Columbia
St Paul’s Hospital
Vancouver V6Z 1Y6, Canada

Introduction

A case-control study is designed to help determine if an exposure is associated with an outcome (i.e., disease or condition of interest). In theory, the case-control study can be described simply. First, identify the cases (a group known to have the outcome) and the controls (a group known to be free of the outcome). Then, look back in time to learn which subjects in each group had the exposure(s), comparing the frequency of the exposure in the case group to the control group.

By definition, a case-control study is always retrospective because it starts with an outcome then traces back to investigate exposures. When the subjects are enrolled in their respective groups, the outcome of each subject is already known by the investigator. This, and not the fact that the investigator usually makes use of previously collected data, is what makes case-control studies ‘retrospective’.

Advantages of Case-Control Studies

Case-control studies have specific advantages compared to other study designs. They are comparatively quick, inexpensive, and easy. They are particularly appropriate for (1) investigating outbreaks, and (2) studying rare diseases or outcomes. An example of (1) would be a study of endophthalmitis following ocular surgery. When an outbreak is in progress, answers must be obtained quickly. An example of (2) would be a study of risk factors for uveal melanoma, or corneal ulcers. Since case-control studies start with people known to have the outcome (rather than starting with a population free of disease and waiting to see who develops it) it is possible to enrol a sufficient number of patients with a rare disease. The practical value of producing rapid results or investigating rare outcomes may outweigh the limitations of case-control studies. Because of their efficiency, they may also be ideal for preliminary investigation of a suspected risk factor for a common condition; conclusions may be used to justify a more costly and time-consuming longitudinal study later.

Cases

Consider a situation in which a large number of cases of post-operative endophthalmitis have occurred in a few weeks. The case group would consist of all those patients at the hospital who developed post-operative endophthalmitis during a pre-defined period. The definition of a case needs to be very specific:

- Within what period of time after operation will the development of endophthalmitis qualify as a case – one day, one week, or one month?

Epidemiology

- Will endophthalmitis have to be proven microbiologically, or will a clinical diagnosis be acceptable?
- Clinical criteria must be identified in great detail. If microbiologic facilities are available, how will patients who have negative cultures be classified?
- How will sterile inflammation be differentiated from endophthalmitis?

There are not necessarily any ‘right’ answers to these questions but they must be answered before the study begins. At the end of the study, the conclusions will be valid only for patients who have the same sort of ‘endophthalmitis’ as in the case definition.

Controls

Controls should be chosen who are similar in many ways to the cases. The factors (e.g., age, sex, time of hospitalisation) chosen to define how controls are to be similar to the cases are the ‘matching criteria’. The selected control group must be at similar risk of developing the outcome; it would not be appropriate to compare a group of controls who had traumatic corneal lacerations with cases who underwent elective intraocular surgery. In our example, controls could be defined as patients who underwent elective intraocular surgery during the same period of time.

Matching Cases and Controls

Although controls must be like the cases in many ways, it is possible to over-match. Over-matching can make it difficult to find enough controls. Also, once a matching variable has been selected, it is not possible to analyse it as a risk factor. Matching for type of intraocular surgery (e.g., secondary IOL implantation) would mean including the same percentage of controls.
Case-Control Studies

as cases who had surgery to implant a secondary IOL; if this were done, it would not be possible to analyse secondary IOL implantation as a potential risk factor for endophthalmitis.

An important technique for adding power to a study is to enrol more than one control for every case. For statistical reasons, however, there is little gained by including more than two controls per case.

Collecting Data
After clearly defining cases and controls, decide on data to be collected; the same data must be collected in the same way from both groups. Care must be taken to be objective in the search for past risk factors, especially since the outcome is already known, or the study may suffer from researcher bias. Although it may not always be possible, it is important to try to mask the outcome from the person who is collecting risk factor information or interviewing patients. Sometimes it will be necessary to interview patients about potential factors (such as history of smoking, diet, use of traditional eye medicines, etc.) in their past. It may be difficult for some people to recall all these details accurately. Furthermore, patients who have the outcome (cases) are likely to scrutinize the past, remembering details of negative exposures more clearly than controls. This is known as recall bias. Anything the researcher can do to minimize this type of bias will strengthen the study.

Analysis; Odds Ratios and Confidence Intervals
In the analysis stage, calculate the frequency of each of the measured variables in each of the two groups. As a measure of the strength of the association between an exposure and the outcome, case-control studies yield the odds ratio. An odds ratio is the ratio of the odds of an exposure in the case group to the odds of an exposure in the control group. It is important to calculate a confidence interval for each odds ratio. A confidence interval that includes 1.0 means that the association between the exposure and outcome could have been found by chance alone and that the association is not statistically significant. An odds ratio without a confidence interval is not very meaningful. These calculations are usually made with computer programmes (e.g., Epi-Info). Case-control studies cannot provide any information about the incidence or prevalence of a disease because no measurements are made in a population based sample.

Risk Factors and Sampling
Another use for case-control studies is investigating risk factors for a rare disease, such as uveal melanoma. In this example, cases might be recruited by using hospital records. Patients who present to hospital, however, may not be representative of the population who get melanoma. If, for example, women present less commonly at hospital, bias might occur in the selection of cases. The selection of a proper control group may pose problems. A frequent source of controls is patients from the same hospital who do not have the outcome. However, hospitalised patients often do not represent the general population; they are likely to suffer health problems and they have access to the health care system. An alternative may be to enroll community controls, people from the same neighborhoods as the cases. Care must be taken with sampling to ensure that the controls represent a normal risk profile. Sometimes researchers enroll multiple control groups. These could include a set of community controls and a set of hospital controls.

Confounders
Matching controls to cases will mitigate the effects of confounders. A confounding variable is one which is associated with the exposure and is a cause of the outcome. If exposure to toxin ‘X’ is associated with melanoma, but exposure to toxin ‘X’ is also associated with exposure to sunlight (assuming that sunlight is a risk factor for melanoma), then sunlight is a potential confounder of the association between toxin ‘X’ and melanoma.

Comment
Case-control studies may prove an association but they do not demonstrate causation.

Consider a case-control study intended to establish an association between the use of traditional eye medicines (TEM) and corneal ulcers. TEM might cause corneal ulcers but it is also possible that the presence of a corneal ulcer leads some people to use TEM. The temporal relationship between the supposed cause and effect cannot be determined by a case-control study.

Be aware that the term ‘case-control study’ is frequently misused. All studies which contain ‘cases’ and ‘controls’ are not case-control studies. One may start with a group of people with a known exposure and a comparison group (‘control group’) without the exposure and follow them through time to see what outcomes result, but this does not constitute a case-control study.

Case-control studies are sometimes less valued for being retrospective. However, they can be a very efficient way of identifying an association between an exposure and an outcome. Sometimes they are the only ethical way to investigate an association. If care is taken with definitions, selection of controls, and reducing the potential for bias, case-control studies can generate valuable information.

Recommended Reading
5. For information about Epi Info (Version 6), a word processing, database, and statistics program for epidemiology on microcomputers, please contact Centers for Disease Control and Prevention, Atlanta, GA 30333 [contact The Division of Surveillance & Epidemiology, Epidemiology Program Office].

Table: Case-Control Studies: Advantages and Disadvantages

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• can obtain findings quickly</td>
<td>• cannot generate incidence data</td>
</tr>
<tr>
<td>• can often be undertaken with minimal funding</td>
<td>• subject to bias</td>
</tr>
<tr>
<td>• efficient for rare diseases</td>
<td>• difficult if record keeping is either inadequate or unreliable</td>
</tr>
<tr>
<td>• can study multiple exposures</td>
<td>• selection of controls can be difficult</td>
</tr>
<tr>
<td>• generally requires few study subjects</td>
<td></td>
</tr>
</tbody>
</table>