Case-Control Studies

Anna Cass, MPH
Greenville Hospital System University Medical Center

Case-Control Studies

- Overview
- When the case-control study design is appropriate
- Benefits
- Potential biases
- Variations on the theme
- Matching
- Analysis
- Study example

The case-control study design can be a time- and cost-efficient choice for a research study. If you are considering a case-control study for your research project, the following slides may help you think through some of the issues relevant to this particular study design.

Case-control studies: The basics

- Retrospective
- Two groups, selected by disease/event status (diseased: cases, non-diseased: control)
- Cases and controls are compared for differences in the prevalence of exposures/factors of interest

The case-control study is an example of a retrospective, analytic study design. It is an observational – rather than interventional – study design. The selection of subjects takes place at the time of or after incidence of the disease (or event of interest), and data are collected to look back toward exposure to risk factors or possible causes of the event.

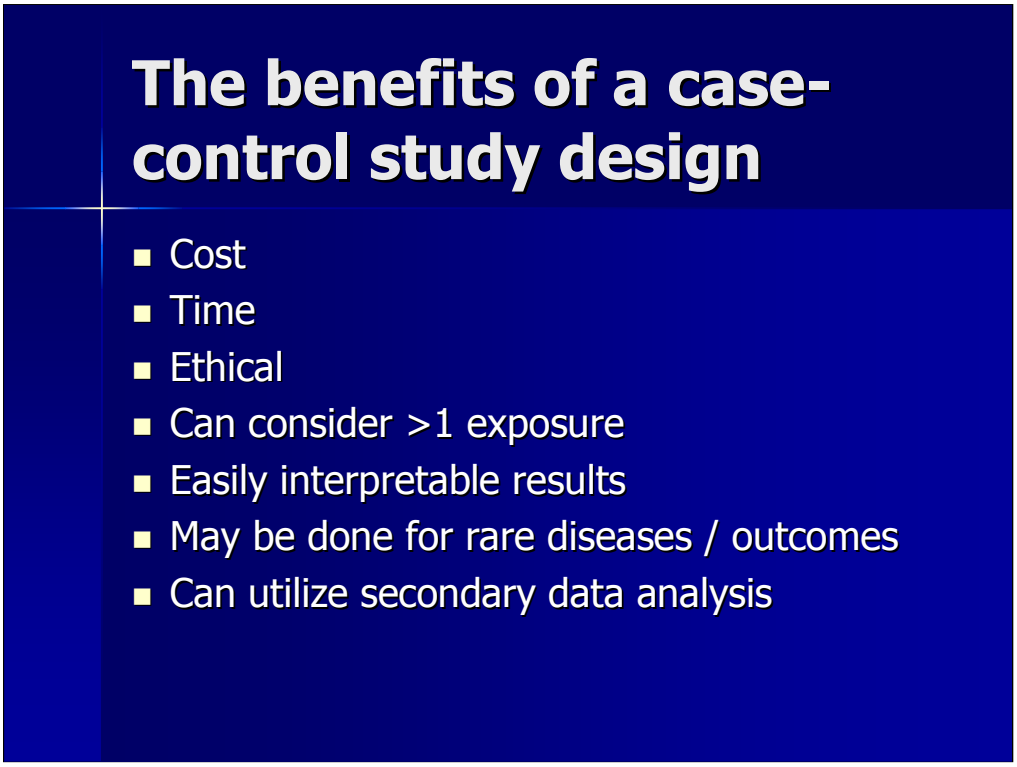
Subjects are selected based on their disease status as either cases (those with the disease or event of interest) or controls (those without the disease or event of interest). Information regarding exposure is collected and can be evaluated in terms of presence or absence of exposure or in terms of level of exposure. Exposure factors can be external (an environmental exposure) or internal (a subject's genetic makeup).

The general idea behind case-control studies is that the prevalence of exposure among cases and controls is compared to determine if the odds of exposure for the two groups differ.

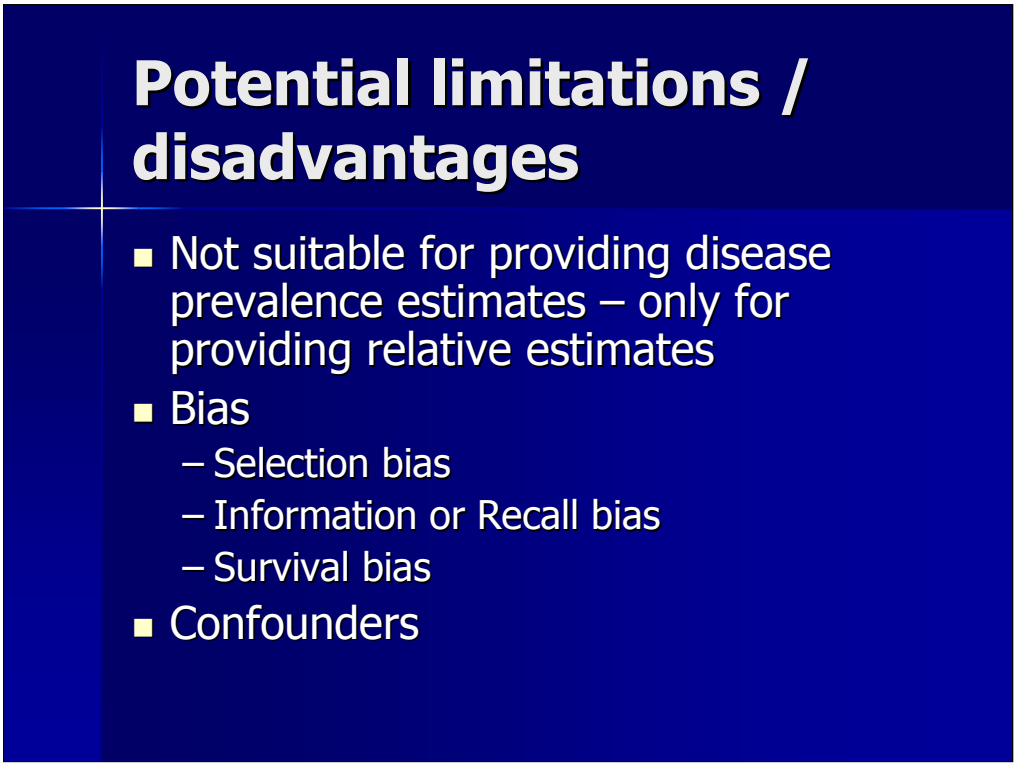
When is a case-control design appropriate?

- Rare outcome? Yes
- Rare exposure? No
- Extended time lapse between exposure and outcome? Yes (with caveats)

Case-control studies are the design of choice for rare outcomes. The case-control design can be an efficient design choice for outcomes in which there may be an extended amount of time between the exposure and the outcome. When this is the case, however, there are potential bias issues to consider if one is relying on subject recall to capture exposure information. The case-control design is not efficient if the exposure of interest is a rare one.



The case-control study design can be attractive due to its speed and efficiency, especially in comparison with prospective study designs. Follow-up time is not needed, as the outcome has already occurred at the time of subject recruitment. The case-control study design is often considered a safe choice ethically, as the researcher is not intervening or affecting the exposure, which has occurred previous to subject involvement in the study. Secondary data analysis, that is, using data that was originally collected for another purpose, can often be incorporated into the study design and increases both the cost-effectiveness and time-efficiency of the study. Further, case-control studies provide easily interpretable results; people are familiar with odds ratios.



The analysis corresponding to case-control studies is designed to provide relative estimates of disease risk, that is, the risk of exposure for cases as compared to controls. Case-control studies (with the possible exception of population-based case-control studies) are not designed to provide absolute prevalence or incidence rates of diseases.

Case-control studies can be particularly susceptible to certain types of biases as well as confounding, and it is important to consider these potential limitations and make accommodations where possible to minimize them through proper population selection and survey design.

Consider the following types of bias and their implications.

One potential type of bias is selection bias. When selecting a study population, controls should be representative of the cases. Controls should also be representative of the general population in terms of their probability of exposure. If the controls are artificially more like the cases, then your results will underestimate the true odds ratio. If the controls are artificially less like the cases, then your resulting odds ratio will be overestimated. It is important to think through these potential differences and consider in what ways they may differ and what the most appropriate source pool would be.

Recall bias is a common concern for case-control studies which utilize surveys to collect study data. Cases and controls may remember or report their exposures differently. There is a tendency for cases (or the survey respondent related to the case) to over-report exposures relative to controls (or the respondents for controls). If cases systematically recall exposures more extensively than controls, the study results will be biases away from the null hypothesis, and the effect of the exposure may be overestimated.

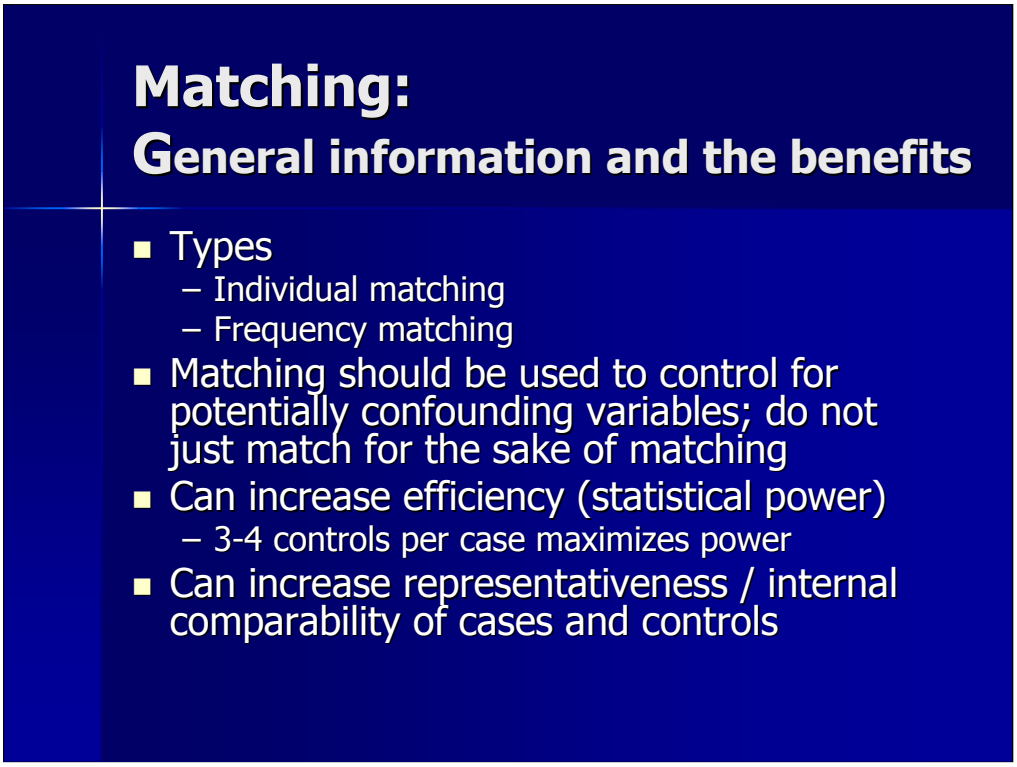
Especially when studying diseases or events with a relatively high mortality rate, it is important to consider whether it is appropriate to assume that cases surviving to the time of study recruitment are representative of all cases in terms of exposure. Think of what has occurred between the time of exposure and/or disease and the time of recruitment. Are there losses among cases or controls, and are those “lost” representative of the others in regard to exposure status? Concurrent recruitment, in which cases are recruited as they are diagnosed, can be used to minimize survival bias.

Case-control studies, like many other study designs, are subject to confounding – that is, the effect of factors that are associated with both the exposure and the outcome. It is important to consider the most appropriate way to address confounding, which can occur either during the design or analysis phase of the study.

Variations of the case-control design

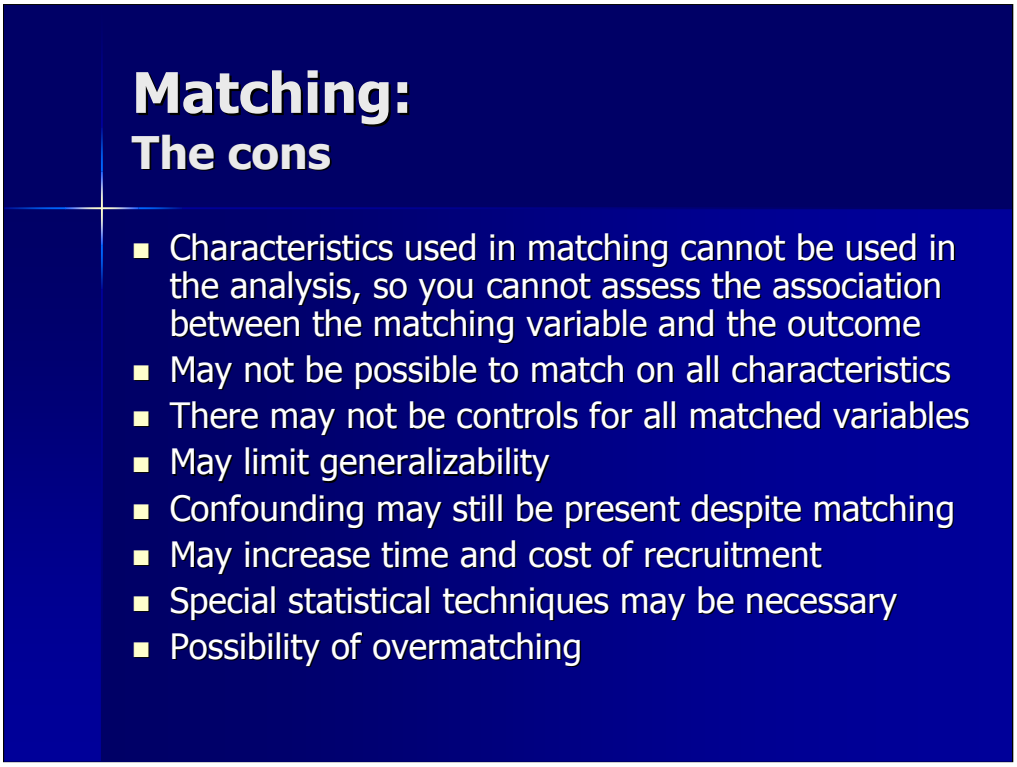
- Prevalent case-control
- Incident case-control
- Nested case-control
- Case-cohort
- Matched case-control

You may want to consider some of the possible variations of the classic case-control study design, each with its own benefits to offer. Nested case-control studies and case-cohort studies occur within a defined cohort. The selection of controls for a nested case-control design occurs at the time of occurrence of each case. For a case-cohort design, controls are randomly selected from the original cohort. In this situation, it is possible for an individual to be both a case and a control.



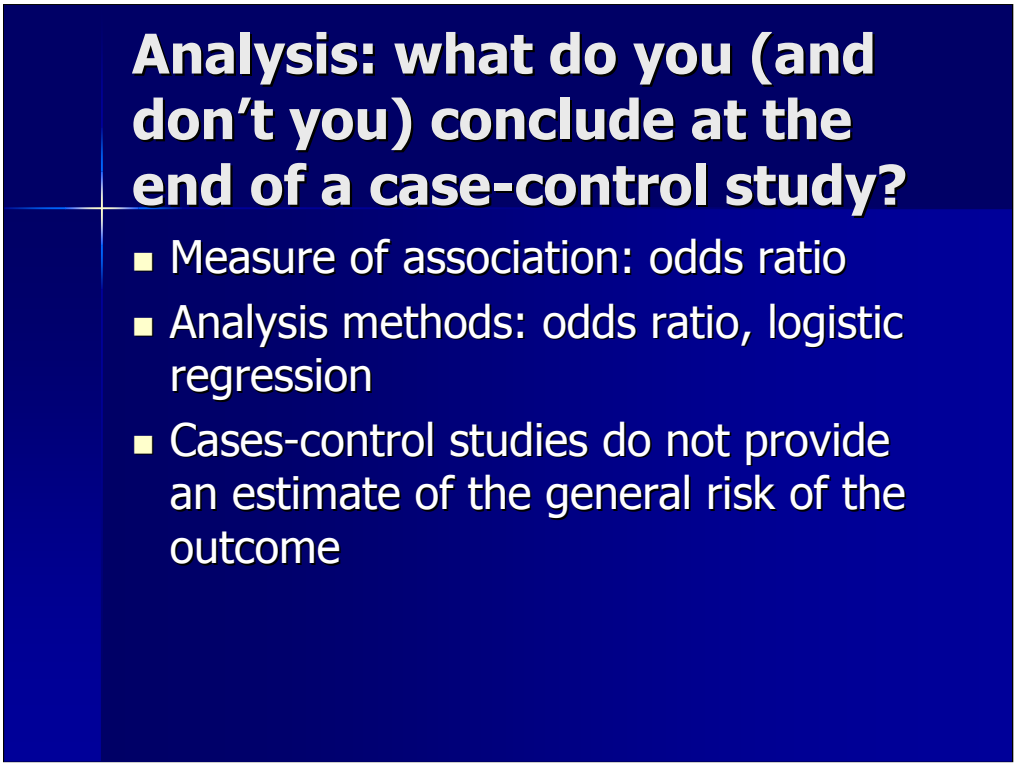
Matching is a method commonly used in case-control studies to address confounding and increase efficiency of the study. In matching, controls are chosen with specific characteristics that correspond to characteristics of cases. This can be done on an individual basis (e.g., for every 25-30 year-old female case, a 25-30 year-old female control is chosen) or based on the overall distribution of characteristics among the cases (if cases are 75% female, then 75% of controls chosen are female).

Matching can be an effective way to control confounding in a case-control study, and should be used for characteristics that are known or strongly suspected to be confounders.



Matching has many benefits, but should be considered carefully, as it cannot be undone later in the study. Variables chosen for matching cannot be assessed for their association with the outcome of interest. Matching should be approached carefully, for if done inappropriately can result in difficulty in recruitment of controls, limited generalizability, residual confounding, and/or decreased statistical efficiency.

See Szklo & Nieto, Epidemiology: Beyond the Basics (2004), pages 43-47 for a thorough explanation of matching, including a list of pros and cons for matching.



Case-control studies are meant to establish a relationship between an exposure and an outcome and are not designed to produce disease incidence or prevalence estimates nor an absolute disease risk/probability.

The measure of association that is the result of a case-control study is an exposure odds ratio, in which the odds of exposure for cases are compared to the odds of exposure for controls. An exposure odds ratio of 5 would be interpreted in the following way: the odds of exposure among cases was five times the odds of exposure among controls. Odds ratios can be calculated from a simple 2x2 table using case/control status and exposure status or from a logistic regression analysis.

Logistic regression is often used in case-control studies to examine the association between the dichotomous outcome (case/control) and one or more exposure factors. This type of analysis can consider the primary exposure factor of the hypothesis while controlling for other factors or patient characteristics. Logistic regression produces an odds ratio as the measure of association.

The odds ratio may approximate the relative risk for rare diseases. Additionally, the frequency of exposure among cases must be representative of cases in the population, and the frequency of exposure among the controls must be representative of controls in the population in order for the odds ratio to be a good estimate of the relative risk.

Sir Richard Doll: A Landmark Case-Control Study

- Doll R, Hill AB. Smoking and carcinoma of the lung: preliminary report. BMJ 1950;221: 739-48.
- Cases: hospitalized patients with lung cancer
- Controls: hospitalized patients with illnesses other than lung cancer
- Exposure: cigarette smoking (included factors of length of time as a smoker and amount smoked)
- Methods: concurrent recruitment; controls matched on hospital, age, & sex; subjects surveyed about smoking history
- Conclusions: "The risk of developing the disease increases in proportion to the amount smoked. It may be 50 times as great among those who smoke 25 or more cigarettes a day as among non-smokers."

In 1950, Doll and Hill published their findings from a case-control study investigating carcinoma of the lung. This landmark study found a strong association between lung cancer and cigarette smoking and was the impetus for the British Doctors cohort study, which continued for decades to provide evidence for the association. Doll and Hill's study also pushed forward the observational study, and specifically the case-control study as an accepted epidemiologic design.

Some interesting commentaries have been written on this study in recent years:

•Schlesselman JJ. The emerging case-control study: Lung cancer in relation to tobacco smoking. Preventive Medicine,. 2006;43(4):251-5.

•Doll SR. Smoking and lung cancer. Am J Respir Crit Care Med. 2000;162(1):4-6.

Resources

- Texts
  - Timmreck, An Introduction to Epidemiology
  - Szklo & Nieto, Epidemiology: Beyond the Basics
- Online videos
  - http://www.videojug.com/interview/epidemiological-study-of-disease-2#what-is-a-case-control-study