Case-Control Studies

Second Edition Authors:

Lorraine K. Alexander, DrPH

Brettania Lopes, MPH

Kristen Ricchetti-Masterson, MSPH

Karin B. Yeatts, PhD, MS

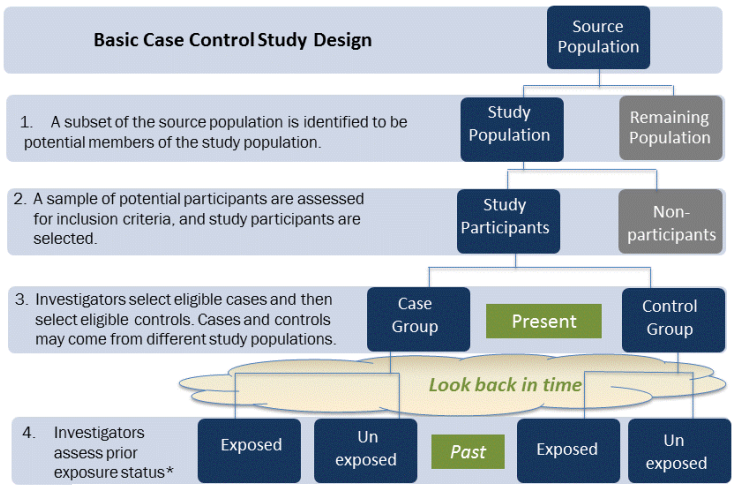
Case-control studies are used to determine if there is an association between an exposure and a specific health outcome. These studies proceed from effect (e.g. health outcome, condition, disease) to cause (exposure). Case-control studies assess whether exposure is disproportionately distributed between the cases and controls, which may indicate that the exposure is a risk factor for the health outcome under study. Case-control studies are frequently used for studying rare health outcomes or diseases.

Unlike cohort or cross-sectional studies, subjects in case-control studies are selected because they have the health outcome of interest (cases). Selection is not based on exposure status. Controls, persons who are free of the health outcome under study, are randomly selected from the population out of which the cases arose. The case-control study aims to achieve the same goals (comparison of exposed and unexposed) as a cohort study but does so more efficiently, by the use of sampling.

At baseline:

- Selection of cases and controls based on health outcome or disease status
- Exposure status is unknown

After cases and controls have been identified, the investigator determines the proportion of cases and the proportion of controls that have been



*Exposure at some specified point before disease onset

exposed to the exposure of interest. Thus, the denominators obtained in a case-control study do not represent the total number of exposed and non-exposed persons in the source population.

After the investigator determines the exposure, a table can be formed from the study data.

	Cases	Controls
Exposed	a	b
Unexposed	c	d

Measures of incidence in case-control studies

In case-control studies the proportion of cases in the entire population-at-risk is unknown, therefore one cannot measure incidence of the health outcome or disease. The controls are representative of the population-at-risk, but are only a sample of that population, therefore the denominator for a risk measure, the population- at-risk, is unknown. We decide on the number of diseased people (cases) and non-diseased people (controls) when we design our study, so the ratios of controls to cases is not biologically or substantively meaningful. However, we can obtain a valid estimate of the risk ratio or rate ratio by using the exposure odds ratio (OR).*

Odds of exposure among cases = a/c

Odds of exposure among controls = b/d

Diseased person-years

	Disease	No Disease
Exposed	a	n_1
Unexposed	c	n_2

RR = (a/n1)/(c/n2)

***Note:** Under some conditions, the odds ratio approximates a risk ratio or rate ratio. However, this is not always the case, and care should be taken to interpret odds ratios appropriately.*

Case-Control Study

	Cases	Controls
Exposed	a	b
Unexposed	c	d

OR = (a/c)/(b/d) = (a/b)/(c/d) = (axd)/(cxb)

If b and d (from the case-control study) are sampled from the source population, n1 + n2, then b will represent the n1 component of the cohort and d will represent the n2 component, and (a/n1)/(c/n2) will be estimated by (a/b)/(c/d).

Interpreting the odds ratio

The odds ratio is interpreted the same way as other ratio measures (risk ratio, rate ratio, etc.).

OR = 1	Odds of disease is the same for exposed and unexposed
OR > 1	Exposure increases odds of disease
OR < 1	Exposure reduces odds of disease

For example, investigators conducted a case-control study to determine if there is an association between colon cancer and a high fat diet. Cases were all confirmed colon cancer cases in North Carolina in 2010. Controls were a sample of North Carolina residents without colon cancer. The odds ratio was 4.0. This odds ratio tells us that individuals who consumed a high fat diet have four times the odds of colon cancer than do individuals who do not consume a high-fat diet. In another study of colon cancer and coffee consumption, the OR was 0.60. Thus, the odds of colon cancer among coffee drinkers is only 0.60 times the odds among individuals who do not consume coffee. This OR tells us that coffee consumption seems to be protective against colon cancer.

Types of case-control studies

Case-control studies can be categorized into different groups based on when the cases develop the health outcome and based on how controls are sampled. Some

case-control studies use prevalent cases while other case-control studies use incident cases. There are also different ways that cases can be identified, such as using population-based cases or hospital-based cases.

Types of cases used in case control studies

Prevalent cases are all persons who were existing cases of the health outcome or disease during the observation period. These studies yield a prevalence odds ratio, which will be influenced by the incidence rate and survival or migration out of the prevalence pool of cases, and thus does not estimate the rate ratio. Case control studies can also use incident cases, which are persons who newly develop the health outcome or disease during the observation period. Recall that prevalence is influenced by both incidence and duration. Researchers that study causes of disease typically prefer incident cases because they are usually interested in factors that lead up to the development of disease rather than factors that affect duration.

Selecting controls

Selection of controls is usually the most difficult part of conducting a case-control study. We will discuss 3 possible ways to select controls:

1. Base or case-base sampling
2. Cumulative density or survivor sampling
3. Incidence density or risk set sampling

Base sampling or case-base sampling

This sampling involves using controls selected from the source population such that every person has the same chance of being included as a control. This type of sampling only works with a previously defined cohort. In these case-control studies, the odds ratio provides a valid estimate of the risk ratio without assuming that the disease is rare in the source population.

Cumulative density sampling or survivor sampling

When controls are sampled from those people who remained free of the health outcome at the end of follow-up then we call the sampling cumulative density sampling or survivor sampling. Controls cannot ever have the outcome (become cases) when using this type of sampling. In these case-control studies, the odds ratio estimates the rate ratio only if the health outcome is rare, i.e. if the proportion of those with the health outcome among each exposure group is less than 10% (requires the rare disease assumption).

Incidence density sampling or risk set sampling

When cases are incident cases and when controls are selected from the at-risk source population at the same time as cases occur (controls must be eligible to become a case if the health outcome develops in the control at a later time during the period of observation) then we call this type of sampling incidence density sampling or risk set sampling. The control series provides an estimate of the proportion of the total person-time for exposed and unexposed cohorts in the source population. In these case -control studies, the odds ratio estimates the rate ratio of cohort studies, without assuming that the disease is rare in the source population.

Note that it is possible, albeit rare, that a control selected at a later time point could become a case during the remaining time that the study is running. This differs from case-control studies that use cumulative density sampling or survivor sampling, which select their controls after the conclusion of the study from among those individuals remaining at risk.

Selecting controls in a risk set sampling or incidence density sampling manner provides two advantages:

1. A direct estimate of the rate ratio is possible.
2. The estimates are not biased by differential loss to follow up among the exposed vs. unexposed controls.

For example, if a large number of smokers left the source population after a certain time point, they would not be available for selection at the end of the study – when controls would be selected in a study that uses cumulative density sampling or survivor sampling. This would give the

investigators biased information regarding the level of exposure among the controls over the course of the study.

Source populations for case-control studies

Source populations can be restricted to a population of particular interest, e.g. postmenopausal women at risk of breast cancer. This restriction makes it easier to control for extraneous confounders in the population. Controls should represent the restricted source population from which cases arise, not all non-cases in the total population. The cases in the study do not have to include all cases in the total population.

Sources of cases

- Cases diagnosed in a hospital or clinic
- Cases entered into a disease registry, e.g. cancer, birth defects, deaths
- Cases identified through mass screening, e.g. hypertensives, diabetics
- Cases identified through a prior cohort study, e.g. lung cancers in an occupational asbestos cohort

Sources of controls

- Population controls are non-cases sampled from the source population giving rise to cases. This is the most desirable method for selecting controls. Sampling randomly from census block groups, or a registry such as the Department of Motor Vehicles (of adults who are able to drive) are examples of ways to find and recruit population-based controls.
- Neighborhood or friend controls are appropriate for selection as controls if these individuals would be included as cases if they developed the health outcome of interest. It is not appropriate to select neighbors or friends as controls if they share the exposure of interest.
- Hospital controls - There are certain problems with hospital controls in that they may not be from the same source population from which the cases arose. Hospital controls may not be representative of the exposure prevalence in the source population of cases, e.g. there may be a higher prevalence of smokers in hospitals. Hospital controls also may have diseases resulting from the exposure of interest, e.g. the exposure (smoking) is related to the disease of interest (cancer) and to heart and lung diseases from which the controls may be suffering.
- Controls with another disease - However if the study is on lung cancer, for example, it is essential to exclude cancers known or suspected to be related to the study exposure of interest. These controls also share some of the same problems as hospital controls.

Advantages of case-control studies

Case-control studies are the most efficient design for rare diseases and require a much smaller study sample than cohort studies. Additionally, investigators can avoid the logistical challenges of following a large sample over time. Thus, case-control studies also allow more intensive evaluation of exposures of cases and controls. Case-control studies that use incidence density sampling or risk set sampling yield a valid estimate of the rate ratio derived from a cohort study if incident cases are studied and controls are sampled from the risk set of the source population. If properly performed (i.e. appropriate sampling), case-control studies provide information that mirrors what could be learned from a cohort study, usually at considerably less cost and time.

Disadvantages of case-control studies

Case-control studies do not yield an estimate of rate or risk, as the denominator of these measures is not defined. Case-control studies may be subject to recall bias if exposure is measured by interviews and if recall of exposure differs between cases and controls. However, investigators may be able to avoid this problem if historical records are available to assess exposure. Choosing an appropriate source population is also difficult and may contribute to selection bias. Case-control studies are not an efficient means for studying rare exposures (less than 10% of controls are exposed) because very large numbers of cases and controls are needed to detect the effects of rare exposures.

Terminology

Cohort studies: An observational study in which subjects are sampled based on the presence (exposed) or absence(unexposed) of a risk factor of interest. These subjects are followed over time for the development of a health outcome of interest.

Cross-sectional studies: An observational study in which subjects are sampled at one point in time, and then the associations between the concurrent risk factors and health outcomes are investigated.

Exposure odds ratio (OR): the odds of a particular exposure among persons with a specific health outcome divided by the corresponding odds of exposure among persons without the health outcome of interest. Yields a valid estimate of the incidence rate ratio or risk ratio derived from a cohort study, depending on control sampling.

Incident case: a person who is newly diagnosed as a case.

Prevalent case: a person who has a health outcome of interest that was diagnosed in the past.

Risk ratio (RR): the likelihood of a particular health outcome occurrence among persons exposed to a given risk factor divided by the corresponding likelihood among unexposed persons.

Source population: the population out of which the cases arose.

From: Medical Epidemiology, R.S. Greenberg, 1993, 1996.

Practice Questions

Answers are located at the end of this notebook

1) Researchers conduct a case-control study of breast cancer, using incident cases. The researchers find out that 90% of the cases had taken hormonal contraceptives in the past. Should the researchers conclude that hormonal contraceptives increase the risk of developing breast cancer?

2) Researchers conduct a case-control study of pancreatic cancer. The study included 200 cases and 200 controls. Of the cases, 80% reported they smoked cigarettes. Among the controls, 50% reported they smoked cigarettes.

a) Prepare a 2x2 table with these data

b) Calculate the exposure odds ratio

c) Interpret the exposure odds ratio in a sentence

References

Dr. Carl M. Shy, Epidemiology 160/600 Introduction to Epidemiology for Public Health course lectures, 1994-2001, The University of North Carolina at Chapel Hill, Department of Epidemiology

Rothman KJ, Greenland S. Modern Epidemiology. Second Edition. Philadelphia: Lippincott Williams and Wilkins, 1998.

The University of North Carolina at Chapel Hill, Department of Epidemiology Courses: Epidemiology 710, Fundamentals of Epidemiology course lectures, 2009-2013, and Epidemiology 718, Epidemiologic Analysis of Binary Data course lectures, 2009-.2013.

Acknowledgement

The authors of the Second Edition of the ERIC Notebook would like to acknowledge the authors of the ERIC Notebook, First Edition: Michel Ibrahim, MD, PhD, Lorraine Alexander, DrPH, Carl Shy, MD, DrPH, and Sherry Farr, GRA, Department of Epidemiology at the University of North Carolina at Chapel Hill. The First Edition of the ERIC Notebook was produced by the Educational Arm of the Epidemiologic Research and Information Center at Durham, NC. The funding for the ERIC Notebook First Edition was provided by the Department of Veterans Affairs (DVA), Veterans Health Administration (VHA), Cooperative Studies Program (CSP) to promote the strategic growth of the epidemiologic capacity of the DVA.

Answers to Practice Questions

1. No, The information provided in this question is only for cases. No information was given in the question about the control group that was studied. To assess if exposure to hormonal contraceptives was associated with the risk of breast cancer, information would be required on the proportion of control subjects who had taken hormonal contraceptives and the researchers would need to calculate the exposure odds ratio.

2.

a) 2x2 table

	Cases	Controls	
Exposed	160 (a)	100 (b)	260
Unexposed	40 (c)	100 (d)	140
	200	200	400

b) Exposure odds ratio = (a/c) / (b/d) = (a*d)/(c*b) = (160*100) / (40*100) = 4.0

c) An odds ratio of 4.0 means that the odds of smokers being a case are 4 times the odds of non-smokers being a case.