

Methodology Series Module 2: Case-control Studies

Maninder Singh Setia

Abstract

Case-Control study design is a type of observational study. In this design, participants are selected for the study based on their outcome status. Thus, some participants have the outcome of interest (referred to as cases), whereas others do not have the outcome of interest (referred to as controls). The investigator then assesses the exposure in both these groups. The investigator should define the cases as specifically as possible. Sometimes, definition of a disease may be based on multiple criteria; thus, all these points should be explicitly stated in case definition. An important aspect of selecting a control is that they should be from the same 'study base' as that of the cases. We can select controls from a variety of groups. Some of them are: General population; relatives or friends; and hospital patients. Matching is often used in case-control control studies to ensure that the cases and controls are similar in certain characteristics, and it is a useful technique to increase the efficiency of the study. Case-Control studies can usually be conducted relatively faster and are inexpensive – particularly when compared with cohort studies (prospective). It is useful to study rare outcomes and outcomes with long latent periods. This design is not very useful to study rare exposures. Furthermore, they may also be prone to certain biases – selection bias and recall bias.

Epidemiologist, MGM Institute of Health Sciences, Navi Mumbai, Maharashtra, India

Address for correspondence:

Dr. Maninder Singh Setia,
MGM Institute of Health Sciences,
Navi Mumbai, Maharashtra, India.
E-mail: maninder.setia@
karanamconsultancy.in

Key Words: Case-control studies, design, limitations, strengths

Introduction

Case-Control study design is a type of observational study design. In an observational study, the investigator does not alter the exposure status. The investigator measures the exposure and outcome in study participants, and studies their association.

participants have the outcome (cases) and others do not have the outcome (controls). During the study procedures, we will examine the exposure of interest in cases as well as controls. We will then study the association between the exposure and outcome in these study participants.

Design

In a case-control study, participants are selected for the study based on their outcome status. Thus, some participants have the outcome of interest (referred to as cases), whereas others do not have the outcome of interest (referred to as controls). The investigator then assesses the exposure in both these groups. Thus, by design, in a case-control study the outcome has to occur in some of the participants that have been included in the study.

As seen in Figure 1, at the time of entry into the study (sampling of participants), some of the study

Examples of Case-control Studies

Smoking and lung cancer study

In their landmark study, Doll and Hill (1950) evaluated the association between smoking and lung cancer. They included 709 patients of lung carcinoma (defined as cases). They also included 709 controls from general medical and surgical patients. The selected controls were similar to the cases with respect to age and sex. Thus, they included 649 males and 60 females in cases as well as controls.

They found that only 0.3% of males were non-smokers among cases. However, the proportion of non-smokers

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

How to cite this article: Setia MS. Methodology series module 2: Case-control studies. Indian J Dermatol 2016;61:146-51.

Received: January, 2016. **Accepted:** January, 2016.

Access this article online

Quick Response Code:



Website: www.e-ijd.org

DOI: 10.4103/0019-5154.177773

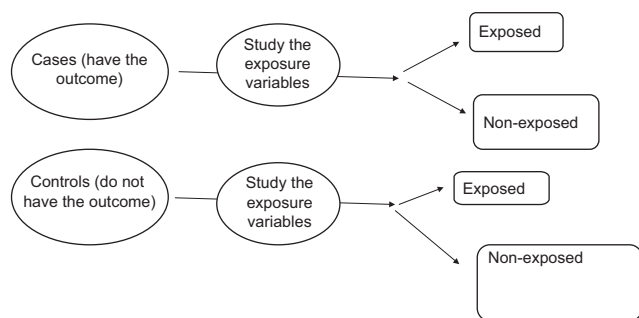


Figure 1: Example of a case-control study

among controls was 4.2%; the difference was statistically significant ($P = 0.00000064$). Similarly they found that about 31.7% of the female were non-smokers in cases compared with 53.3% in controls; this difference was also statistically significant ($0.01 < p < 0.02$).

Melanoma and tanning (Lazovic et al., 2010)

The authors conducted a case-control study to study the association between melanoma and tanning. The 1167 cases - individuals with invasive cutaneous melanoma - were selected from Minnesota Cancer Surveillance System. The 1101 controls were selected randomly from Minnesota State Driver's License list; they were matched for age (± 5 years) and sex.

The data were collected by self administered questionnaires and telephone interviews. The investigators assessed the use of tanning devices (using photographs), number of years, and frequency of use of these devices. They also collected information on other variables (such as sun exposure; presence of freckles and moles; and colour of skin, hair, among other exposures).

They found that melanoma was higher in individuals who used UVB enhances and primarily UVA-emitting devices. The risk of melanoma also increased with increase in years of use, hours of use, and sessions.

Risk factors for erysipelas (Pitché et al, 2015)

Pitché et al (2015) conducted a case-control study to assess the factors associated with leg erysipelas in sub-Saharan Africa. This was a multi-centre study; the cases and controls were recruited from eight countries in sub-Saharan Africa.

They recruited cases of acute leg cellulitis in these eight countries. They recruited two controls for each case; these were matched for age (± 5 years) and sex. Thus, the final study has 364 cases and 728 controls. They found that leg erysipelas was associated with obesity, lymphoedema, neglected traumatic wound, toe-web intertrigo, and voluntary cosmetic depigmentation.

We have provided details of all the three studies in the bibliography. We strongly encourage the readers to read the papers to understand some practical aspects of case-control studies.

Selection of Cases and Controls

Selection of cases and controls is an important part of this design. Wacholder and colleagues (1992 a, b, and c) have published wonderful manuscripts on design and conduct of case-control of studies in the American Journal of Epidemiology. The discussion in the next few sections is based on these manuscripts.

Selection of case

The investigator should define the cases as specifically as possible. Sometimes, definition of a disease may be based on multiple criteria; thus, all these points should be explicitly stated in case definition.

For example, in the above mentioned Melanoma and Tanning study, the researchers defined their population as any histologic variety of invasive cutaneous melanoma. However, they added another important criterion - these individuals should have a driver's license or State identity card. This probably is not directly related to the clinic condition, so why did they add this criterion? We will discuss this in detail in the next few paragraphs.

Selection of a control

The next important point in designing a case-control study is the selection of control patients.

In fact, Wacholder and colleagues have extensively discussed aspects of design of case control studies and selection of controls in their article.

According to them, an important aspect of selecting a control is that they should be from the same 'study base' as that of the cases. Thus, the pool of population from which the cases and controls will be enrolled should be same. For instance, in the Tanning and Melanoma study, the researchers recruited cases from Minnesota Cancer Surveillance System; however, it was also required that these cases should either have a State identity card or Driver's license. This was important since controls were randomly selected from Minnesota State Driver's license list (this also included the list of individuals who have the State identity card).

Another important aspect of a case-control study is that we should measure the exposure similarly in cases and controls. For instance, if we design a research protocol to study the association between metabolic syndrome (exposure) and psoriasis (outcome), we should ensure that we use the same criteria (clinically and biochemically) for evaluating metabolic syndrome in cases and controls. If we use different criteria to measure the metabolic syndrome, then it may cause information bias.

Types of Controls

We can select controls from a variety of groups. Some of them are: General population; relatives or friends; or hospital patients.

Hospital controls

An important source of controls is patients attending the hospital for diseases other than the outcome of interest. These controls are easy to recruit and are more likely to have similar quality of medical records.

However, we have to be careful while recruiting these controls. In the above example of metabolic syndrome and psoriasis, we recruit psoriasis patients from the Dermatology department of the hospital as controls. We recruit patients who do not have psoriasis and present to the Dermatology as controls. Some of these individuals have presented to the Dermatology department with tinea pedis. Do we recruit these individuals as controls for the study? What is the problem if we recruit these patients? Some studies have suggested that diabetes mellitus and obesity are predisposing factors for tinea pedis. As we know, fasting plasma glucose of >100 mg/dl and raised triglycerides (≥ 150 mg/dl) are criteria for diagnosis of metabolic syndrome. Thus, it is quite likely that if we recruit many of these tinea pedis patients, the exposure of interest may turn out to be similar in cases and controls; this exposure may not reflect the truth in the population.

Relative and friend controls

Relative controls are relatively easy to recruit. They can be particularly useful when we are interested in trying to ensure that some of the measurable and non-measurable confounders are relatively equally distributed in cases and controls (such as home environment, socio-economic status, or genetic factors).

Another source of controls is a list of friends referred by the cases. These controls are easy to recruit and they are also more likely to be similar to the cases in socio-economic status and other demographic factors. However, they are also more likely to have similar behaviours (alcohol use, smoking etc.); thus, it may not be prudent to use these as controls if we want to study the effect of these exposures on the outcome.

Population controls

These controls can be easily conducted the list of all individuals is available. For example, list from state identity cards, voter's registration list, etc., In the Tanning and melanoma study, the researchers used population controls. They were identified from Minnesota state driver's list.

We may have to use sampling methods (such as random digit dialing or multistage sampling methods) to recruit controls from the population. A main advantage is

that these controls are likely to satisfy the 'study-base' principle (described above) as suggested by Wacholder and colleagues. However, they can be expensive and time consuming. Furthermore, many of these controls will not be inclined to participate in the study; thus, the response rate may be very low.

Matching in a Case-control Study

Matching is often used in case-control control studies to ensure that the cases and controls are similar in certain characteristics. For example, in the smoking and lung cancer study, the authors selected controls that were similar in age and sex to carcinoma cases. Matching is a useful technique to increase the efficiency of study.

'Individual matching' is one common technique used in case-control study. For example, in the above mentioned metabolic syndrome and psoriasis, we can decide that for each case enrolled in the study, we will enroll a control that is matched for sex and age (± 2 years). Thus, if 40 year male patient with psoriasis is enrolled for the study as a case, we will enroll a 38-42 year male patient without psoriasis (and who will not be excluded for other reason) as controls.

If the study has used 'individual matching' procedures, then the data should also reflect the same. For instance, if you have 45 males among cases, you should also have 45 males among controls. If you show 60 males among controls, you should explain the discrepancy.

Even though matching is used to increase the efficiency in case-control studies, it may have its own problems. It may be difficult to find the exact matching control for the study; we may have to screen many potential enrollees before we are able to recruit one control for each case recruited. Thus, it may increase the time and cost of the study.

Nonetheless, matching may be useful to control for certain types of confounders. For instance, environment variables may be accounted for by matching controls for neighbourhood or area of residence. Household environment and genetic factors may be accounted for by enrolling siblings as controls.

If we use controls from the past (time period when cases did not occur), then the controls are sometimes referred to historic controls. Such controls may be recruited from past hospital records.

Strengths of a Case-control Study

- Case-Control studies can usually be conducted relatively faster and are inexpensive – particularly when compared with cohort studies (prospective)
- It is useful to study rare outcomes and outcomes with long latent periods. For example, if we wish to study the factors associated with melanoma in India,

it will be useful to conduct a case-control study. We will recruit cases of melanoma as cases in one study site or multiple study sites. If we were to conduct a cohort study for this research question, we may to have follow individuals (with the exposure under study) for many years before the occurrence of the outcome

- It is also useful to study multiple exposures in the same outcome. For example, in the metabolic syndrome and psoriasis study, we can study other factors such as Vitamin D levels or genetic markers
- Case-control studies are useful to study the association of risk factors and outcomes in outbreak investigations. For instance, Freeman and colleagues (2015) in a study published in 2015 conducted a case-control study to evaluate the role of proton pump inhibitors in an outbreak of non-typhoidal salmonellosis.

Limitations of a Case-control Study

- The design, in general, is not useful to study rare exposures. It may be prudent to conduct a cohort study for rare exposures
- We are not able to estimate the incidence or prevalence in a case-control study

Why can't we comment on the incidence or prevalence of the disease?

Since the investigator chooses the number of cases and controls, the proportion of cases may not be representative of the proportion in the population. For instance if we choose 50 cases of psoriasis and 50 controls, the prevalence of proportion of psoriasis cases in our study will be 50%. This is not true prevalence. If we had chosen 50 cases of psoriasis and 100 controls, then the proportion of the cases will be 33%.

- The design is not useful to study multiple outcomes. Since the cases are selected based on the outcome, we can only study the association between exposures and that particular outcome
- Sometimes the temporality of the exposure and outcome may not be clearly established in case-control studies
- The case-control studies are also prone to certain biases

In general, individuals may not be able to recall all exposures accurately. Furthermore, cases are more likely to remember detailed exposure history (particularly behaviours such as dietary habits) compared with controls – particularly population based controls. Thus, this may lead to recall bias

If the cases and controls are not selected similarly from the study base, then it will lead to selection bias.

Analysis

- **Odds Ratio:** We are able to calculate the odds ratios (OR) from a case-control study. Since we are not able to measure incidence data in case-control study, an odds ratio is a reasonable measure of the relative risk (under some assumptions). Additional details about OR will be discussed in the biostatistics section.

Table : Calculating an Odds Ratio (OR)

	Cases	Controls	Total
Exposed	A	B	A + B
Non-exposed	C	D	C + D
Total	A + C	B + D	A + B + C + D

Odds of the cases being exposed= A/C , Odds of the controls being exposed= B/D , Odds ratio=ratio of the two= A/C divided by $B/D=AD/BC$, Odds ratio= AD/BC

Table : Hypothetical study of metabolic syndrome and psoriasis

	Psoriasis	Controls	Total
Metabolic syndrome	60	60	120
No metabolic syndrome	40	140	180
Total	100	200	300

The OR in the above study is 3.5. Since the OR is greater than 1, the outcome is more likely in those exposed (those who are diagnosed with metabolic syndrome) compared with those who are not exposed (those who do are not diagnosed with metabolic syndrome). However, we will require confidence intervals to comment on further interpretation of the OR (This will be discussed in detail in the biostatistics section).

- **Other analysis:** We can use logistic regression models for multivariate analysis in case-control studies. It is important to note that conditional logistic regressions may be useful for matched case-control studies.

Additional Points in A Case-control Study

How many controls can I have for each case?

The most optimum case-to-control ratio is 1:1. Jewell (2004) has suggested that for a fixed sample size, the chi square test for independence is most powerful if the number of cases is same as the number of controls. However, in many situations we may not be able recruit a large number of cases and it may be easier to recruit more controls for the study. It has been suggested that we can increase the number of controls to increase statistical power (if we have limited number of cases) of the study. If data are available at no extra cost, then we may recruit multiple controls for each case. However, if it is expensive to collect exposure and outcome information from cases and controls, then the optimal ratio is 4 controls: 1 case. It has been argued

that the increase in statistical power may be limited with additional controls (greater than four) compared with the cost involved in recruiting them beyond this ratio.

I have conducted a randomised controlled trial. I have included a group which received the intervention and another group which did not receive the intervention. Can I call this a case-control study?

A randomised controlled trial is an experimental study. In contrast, case-control studies are observational studies. These are two different groups of studies. One should not use the word case-control study for a randomised controlled trial (even though you have a control group in the study). Every study with a control group is not a case-control study. For a study to be classified as a case-control study, the study should be an observational study and the participants should be recruited based on their outcome status (some have the disease and some do not).

Should I call case-control studies prospective or retrospective studies?

In 'The Dictionary of Epidemiology' by Porta (2014), the authors have suggested that even though the term 'retrospective' was used for case-control studies, the study participants are often recruited prospectively. In fact, the study on risk factors for erysipelas (Pitché *et al.*, 2015) was a prospective case case-control study. Thus, it is important to remember that the nature of the study (case-control or cohort) depends on the sampling method. If we sample the study participants based on exposure and move towards the outcome, it is a cohort study. However, if we sample the participants based on the outcome (some with outcome and some do not) and study the exposures in both these groups, it is a case-control study.

Summary

In case-control studies, participants are recruited on the basis of disease status. Thus, some of participants have the outcome of interest (referred to as cases), whereas others do not have the outcome of interest (referred to as controls). The investigator then assesses the exposure in both these groups. Case-control studies are less expensive and quicker to conduct (compared with prospective cohort studies at least). The measure of association in this type of study is an odds ratio. This type of design is useful for rare outcomes and those with long latent periods. However, they may also be prone to certain biases – selection bias and recall bias.

Financial support and sponsorship

Nil.

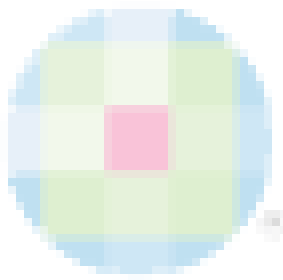
Conflicts of interest

There are no conflicts of interest.

Bibliography

1. Charles H. Hennekens, and Julie E. Buring. (1987). *Epidemiology in Medicine* (First ed.). Philadelphia, USA: Lippincott Williams & Wilkins.
2. Cheng, S., & Chong, L. (2002). A prospective epidemiological study on tinea pedis and onychomycosis in Hong Kong. *Chin Med J (Engl)*, 115 (6), 860-865.
3. Doll, R., & Hill, A. B. (1950). Smoking and carcinoma of the lung; preliminary report. *Br Med J*, 2 (4682), 739-748.
4. Freeman, R., Dabrera, G., Lane, C., Adams, N., Browning, L., Fowler, T., *et al.* (2015). Association between use of proton pump inhibitors and non-typhoidal salmonellosis identified following investigation into an outbreak of Salmonella Mikawasima in the UK, 2013. *Epidemiol Infect*, 1-8.
5. Gail, M., Williams, R., Byar, D. P., & Brown, C. (1976). How many controls? *J Chronic Dis*, 29 (11), 723-731.
6. International Diabetes Foundation. (2006). The IDF consensus worldwide definition of the Metabolic Syndrome. Retrieved 24 January, 2016, from https://www.idf.org/webdata/docs/IDF_Meta_def_final.pdf
7. Jewell N. (2004). *Statistics for Epidemiology*. Boca Raton, US: Chapman and Hall/CRC.
8. Jos W. R. Twisk. (2013). *Applied Longitudinal Data Analysis for Epidemiology* (Second ed.). Cambridge, UK: Cambridge University Press.
9. Kenneth J. Rothman, Sander Greenland, and Timothy L. Lash. (2008). *Modern Epidemiology* (Third ed.). Philadelphia, USA: Lippincott Williams & Wilkins.
10. Kleinbaum D, Kupper L, & Morgenstern H. (1982). *Epidemiologic research*. New York, US: John Wiley and Sons, Inc.
11. Lazovich, D., Vogel, R. I., Berwick, M., Weinstock, M. A., Anderson, K. E., & Warshaw, E. M. (2010). Indoor tanning and risk of melanoma: A case-control study in a highly exposed population. *Cancer Epidemiol Biomarkers Prev*, 19 (6), 1557-1568.
12. Pitché, P., Diatta, B., Faye, O., Diane, B. F., Sangare, A., Niamba, P., *et al.* (2015). [Risk factors associated with leg erysipelas (cellulitis) in sub-Saharan Africa: A multicentre case-control study]. *Ann Dermatol Venereol*, 142 (11), 633-638.
13. Porto Miquel; Assoc. editors, Sander Greenland, Miguel Hernán, Isabel dos Santos Silva, John M. Last; Asst. editor, Andrea Burón. *A Dictionary of Epidemiology* (2014). (Sixth ed.): International Epidemiological Association, Oxford University Press, New York, USA.
14. Song, J. W., & Chung, K. C. (2010). Observational studies: Cohort and case-control studies. *Plast Reconstr Surg*, 126 (6), 2234-2242.
15. Stephen B. Hulley, Steven R. Cummings, Warren S. Browner, Deborah Grady, Norman Hearst, and Thomas B. Newman. (2001). *Designing Clinical Research* (Second ed.). Philadelphia, USA: Lippincott Williams and Wilkins.
16. Szklo M, & Javier Nieto F. (2004). *Epidemiology: Beyond the basics*. Sudbury, MA: Jones and Bartlett Publishers, Inc.
17. Wacholder, S., McLaughlin, J. K., Silverman, D. T., and Mandel, J. S. (1992a). Selection of controls in case-control studies. I. Principles. *Am J Epidemiol*, 135 (9), 1019-1028.

18. Wacholder, S., Silverman, D. T., McLaughlin, J. K., & Mandel, J. S. (1992b). Selection of controls in case-control studies. III. Design options. *Am J Epidemiol*, 135 (9), 1042-1050.
19. Wacholder, S., McLaughlin, J. K., Silverman, D. T., & Mandel, J. S. (1992c). Selection of controls in case-control studies. I. Principles. *Am J Epidemiol*, 135 (9), 1019-1028.
20. Ye, D., Shan, J., Huang, Y., Li, J., Li, C., Liu, X., et al. (2015). A gloves-associated outbreak of imipenem-resistant *Acinetobacter baumannii* in an intensive care unit in Guangdong, China. *BMC Infect Dis*, 15, 179.



Copyright of Indian Journal of Dermatology is the property of Medknow Publications & Media Pvt. Ltd. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.