Column: Methods in Psychiatric Research

CASE-CONTROL STUDIES

Pankajakshan Vijayanthi Indu[1], K Vidhukumar [2]*

[1]Government Medical College, Kozhikode
[2]Government Medical College, Thiruvananthapuram
* Corresponding Address: Additional Professor in Psychiatry, Government Medical College, Thiruvananthapuram-695011, Kerala. Email: kumarv68@gmail.com

ABSTRACT
Case-control studies (CCS) are observational analytic studies done often in instances of rare cases or outcomes. A case-control study compares clearly defined cases and controls arising from the same population for well-measured exposures. The ease of conduct of the study in terms of resources makes CCS popular. However, since the direction of the study is retrospective, selection and measurement biases are potential threats to the validity of conclusions from CCS. Moreover, CCS does not give a direct estimate of the risk. Another issue with the CCS study is confounding. How these validity issues could be addressed are also discussed.
Keywords: Case-control study, selection bias, measurement bias. odds ratio

Case-control studies (CCS) are observational analytical studies in which the association between disease and potential risk factors is assessed by taking samples of cases (with disease) and controls (who are at risk of developing the disease).[1] Then, the frequency of exposure to potential risk factors is measured in both the groups by looking backwards in time. Hence, CCS are also known as retrospective studies.[2] What makes these studies retrospective is that the outcome of each participant is known to the researcher when the subjects are recruited for the study.[3] This study design is used to study the risk factors of rare diseases or outcomes and to investigate outbreaks of acute diseases like infectious diseases. Although risk ratio cannot be computed directly, an estimate of it can be obtained from these studies through odds ratio. [2,3]

An investigator must focus on the following steps while designing a CCS:

- Selection of cases
- Selection of controls
- Measurement of exposure
- Addressing bias
- Addressing confounding

Selection of cases

Cases are those that have developed an outcome of interest. A clear definition of the outcome which is being studied must be provided. This could be clinical symptoms or diagnostic criteria or by using diagnostic tools or laboratory methods. Further, the age range of the participants, the location from which they are selected (hospital-based or population-based) or other eligibility criteria and exclusion criteria have to be explicitly stated.[2] As many sources of information can be used to ascertain the disease status or the outcome of interest like hospital records, death certificates, logbooks, registries, clinical evaluation, laboratory investigations, or diagnostic tools.

A CCS can include all the cases or a representative sample of cases from a defined population. Preferably, new or incident cases, rather than existing prevalent ones, should be selected. If prevalent cases are studied,

Access the article online:
https://kjponline.com/index.php/kjp/article/view/239
DOI: https://doi.org/10.30834/KJP.33.2.2020.239
Received: 1/12/2020. Web publication: 31/12/2020

QR code

Please cite this article as: PV Indu, K Vidhukumar. Case-control studies (Column: Research Methods in Psychiatry). Kerala Journal of Psychiatry 2020;34(1): 84-87

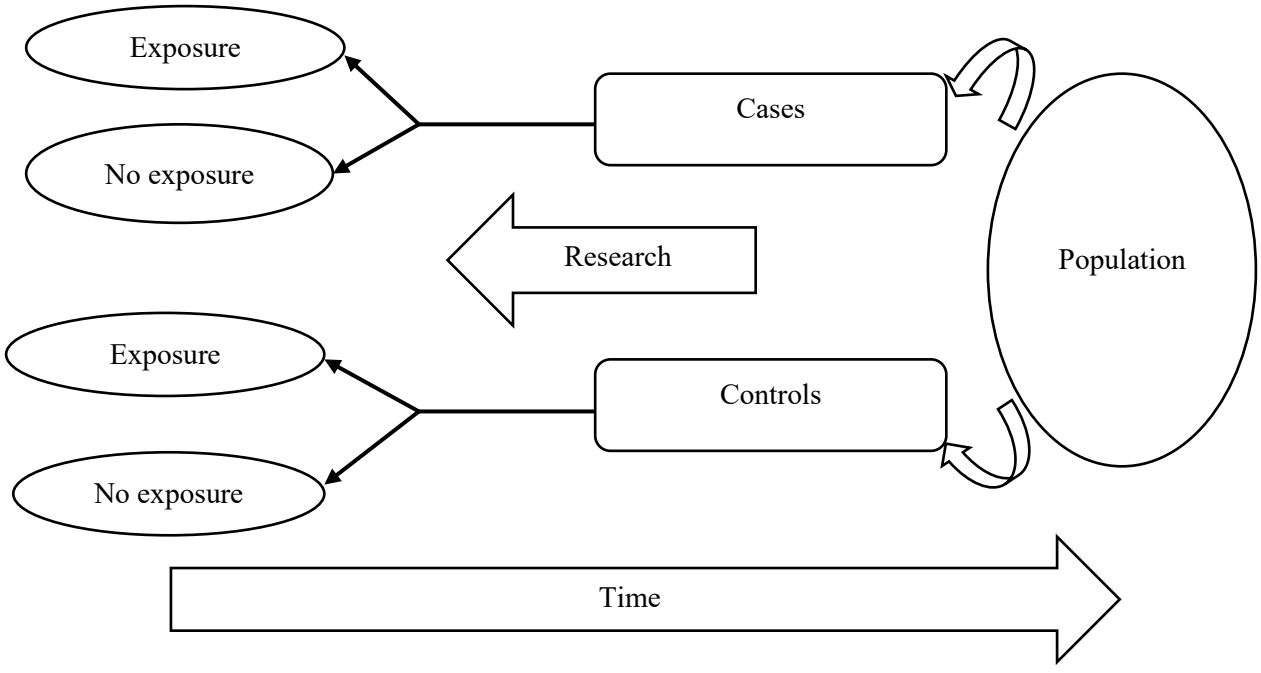

Figure 1. Design of case-control studies

the risk factors identified may be those determining the incidence or development of the disease, duration, or both; the contribution of the exposure to the development of the outcome cannot be determined.[2] So also, incident cases reduce recall bias.[5]

Selection of controls

Cases and controls should be comparable for the case-control study to be valid. For this, the controls would have been cases had they developed the outcome of interest. Cases and controls should be from the same base population and should have an equal chance of being exposed.[2] Controls should not have the disease and should be selected from the same population at risk for the disease; they should represent the target population.[5] The controls provide the background proportion of exposure that is expected in the cases.[4]

The process of selecting controls should be independent of the exposure being studied.[4] Controls can be selected from the population from which the cases were selected based on probability sampling (population-based CCS) or from a cohort from which the cases are drawn (nested CCS). If the cases are selected from a hospital, controls can be selected from patients with different diseases (other than the outcome of interest), from the same setting (hospital-based CCS). Controls can also be identified from the community from which the hospital draws patients. They can also be drawn from among the friends, relatives or neighbours of the cases.[2,6] Multiple control groups can be chosen – one from the same hospital and another from the community from which the cases are drawn. As against this, multiple controls can be selected per case if the number of cases is less or the outcome is rare. This can increase the power of the study. [2,5] A case: control ratio of 1:1 is the most optimum; beyond 4:1, the increase in statistical power is not marked.[7]

Matching the cases and controls for one or more characteristics can be done to ensure that the two groups are comparable. Usually, matching is done for age, sex or place of residence. This is a strategy used to adjust for confounding variables. But matching the cases and controls for too many variables leads to overmatching and can bias the study by finding no significant association.

Measurement of exposure

Exposure information can be collected from the participants or their surrogates (mothers of children or caregivers of dementia patients) through direct interviews, questionnaires, or medical records.

Figure 2. Calculation of Odds Ratio from a case-control study

	Cases	Controls	
Exposure	a	b	a + b
No exposure	c	d	c + d
	a + c	b + d	

Odds of exposure among cases = a/ (a + c) ÷ c/ (a + c) = a/c

Odds of exposure among controls = b/ (b + d) ÷ d/ (b + d) = b/d

Odds ratio = a/c ÷ b/d = ad/bc

interviews, questionnaires, or medical records. Information should be obtained using similar procedures from both cases and controls.[8] There is a greater likelihood of the cases recalling the exposure than the controls, leading to recall bias. The report of exposure by the interviewer can vary systematically between the cases and controls, thereby leading to interviewer bias. Blinding the interviewer to the case-control status and the hypotheses being tested can help reduce this bias.[5] Using multiple sources of information – like treatment records documented before the outcome – to evaluate the exposure can also minimise measurement bias.

Addressing bias

The potential for bias – systematic error in collecting or interpreting data– is a major factor that affects the validity of CCS. This can affect the interpretation of the hypotheses obtained from that data. Selection bias occurs when the exposure of interest affects the selection of cases and controls in some manner,[8] E.g., Heavy smokers refusing to participate in a study to assess the risk of lung cancer can lead to selection bias. Interviewer bias occurs when the interviewer ascertains the exposure with greater enthusiasm from the cases than controls. The observer being aware of the case-control status can contribute to this bias, and blinding can prevent this. Recall bias was alluded to earlier. Both the recall bias and interviewer bias are examples of measurement or ascertainment bias. In the case of a CCS, such a bias leads to misclassification of exposure, thus leading to a wrong estimate of the risk.

A detailed discussion of biases is beyond the scope of this article. The possibility of biases must be considered and addressed carefully while designing a CCS. Appropriate definition and selection of cases and controls, explicit definition and proper and uniform measurement of the study variable and covariates are essential to prevent bias.

Addressing confounding

A confounding variable is independently associated with the exposure and outcome variables. It can distort the effect of the exposure on the outcome.[2] During the designing of the study itself, potential confounders must be identified and addressed. Potential confounders can be addressed in the design phase and the analysis phase of CCS. In the design phase, the potential confounders can be controlled by restriction (exclusion) or matching. However, the effect of these variables on the outcome cannot be assessed further. In the analysis phase, stratification of the data based on the confounding variable and assessment of Mantel Haenszel Odds Ratio, and logistic regression can be used to adjust for the effect of the confounder. For this, planning is required in the design phase to ensure that the data regarding the confounding variable is collected during the study.[4]

Analysis

In CCS, Odds Ratio (OR) is the measure of the strength of association between the exposure and outcome variables. It is the ratio of the odds of exposure among cases to the odds of exposure among controls (See Figure 2). A greater frequency of exposure among cases leads to an OR >1, suggesting that it is a risk factor. On the other hand, a lesser frequency of exposure among cases leads to an OR <1,

suggesting that it is a protective factor. If OR = 1, the odds of exposure is the same in cases and controls. This implies that exposure is neither a risk factor nor a protective factor.[2,6] Univariate analysis provides crude OR, while multivariate analysis provides adjusted OR, adjusted for confounding variables.

It is imperative to assess the 95% Confidence Interval (95% CI) of the OR. If the 95% CI of the OR includes the null value of 1, it can be concluded that the *P*-value from the test of statistical significance would be greater than 0.05. E.g., If the OR is 1.8 with a 95% CI of 0.63 – 4.5, it can be concluded that the association is not statistically significant. On the other hand, if the OR is 1.8, with a 95% CI of 1.2 – 3.4, it can be inferred that the association is statistically significant. Both *P*-value and CI together give maximum information about the role of chance in obtaining the finding.[8]

The OR estimates the relative risk (RR), especially when the incidence of the disease or outcome is low. Generally, when the disease rates in the unexposed population are less than 1/100, the OR becomes approximately equal to the RR. [2]

Advantages of CCS:

- They are quick and inexpensive
- Appropriate design to study rare diseases
- Appropriate design to study diseases with a long latent period
- Multiple exposure factors can be studied for a single outcome. [8]

Disadvantages of CCS:

- Inefficient to study rare exposures
- Cannot study multiple outcomes
- Incidence rates of disease or outcome measures cannot be computed
- The temporal relationship of exposure and outcome is difficult to establish
- Particularly prone for bias like selection and recall bias [7,8]

Conclusion

In designing a CCS, investigators must explicitly define the cases or outcome variable, select controls meticulously and measure exposure variables accurately. The potential biases and confounding factors have to be thought of and addressed in the design phase of the study itself. Properly designed and well-executed CCS can provide ORs which can be reliable estimates of relative risk. They can be very efficient in identifying the association between exposure and outcome variables. While reporting CCS, a checklist of the items to be included is provided by the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) statement.[9] This can aid in providing a thorough description of the methodology used in CCS. A critical appraisal of the methodology and acceptance of the validity of the findings is also made possible.

REFERENCES

1. Breslow NE. Case-Control Studies. In: Ahrens W., Pigeot I. (eds) Handbook of Epidemiology. Springer, Berlin, Heidelberg. 2005. pp. 287-319. Available from: https://doi.org/10.1007/978-3-540-26577-1_7
2. Fletcher RH, Fletcher SW. Risk: Looking backward. In: Clinical epidemiology: The essentials. Fourth Edition. Philadelphia: Lippincott Williams & Wilkins. 2005.
3. Lewallen S, Courtright P. Epidemiology in practice: Case-control studies. Community Eye Health 1998;11(28):57-8.
4. Schulz KF, Grimes DA. Case-control studies: research in reverse. Lancet 2002;359:431-4.
5. Levin KA. Study design V. case-control studies. Evidence-Based Dentistry 2003;7:83-84. doi:10.1038/sj.ebd.6400436
6. Alexander LK, Lopes B, Richetti-Masterson K, Yeatts KB. Case-control studies. ERIC Notebook. Second Edition. Available from: https://sph.unc.edu/wp-content/uploads/sites/112/2015/07/nciph_ERIC5.pdf
7. Sethia MS. Methodology series module: Case-control studies. Indian J Dermatol 2016;61:146-51.
8. Hennekens C, Buring J. Case-control studies. In: Epidemiology in medicine. Boston, Toronto: Little Brown & Co. 1987.
9. von Elm E, Altman DG, Pocock SJ, Gotzsche PC, Vandenbroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007;335:806-8. Available from: doi:10.1136/bmj.39335.541782.ad