Chapter 3
Case–Control Study

Noraini Abdul Ghafar

Learning Objectives

After completing this chapter, you will be able to:

- Describe how case–control study is conducted.
- Understand the issues in selecting controls.
- Describe the advantages and disadvantages of case–control study.
- Explain nested case–control study.
- Calculate the odds ratio with 95% confidence intervals.

1   Introduction

Case–control study is one of the epidemiological study designs that falls under non-experimental or observational study. This kind of study permits the researcher to determine if an exposure is associated with an outcome. Based on the study design, cases (group known to have the outcome or the disease of interest) and controls (group known to be free from the outcome or the disease of interest) will be determined. Researchers then will look back to find the exposure in each group. Each group's exposure level is then assessed according to its prevalence. It is reasonable to deduce that exposure may be linked to either an increased or decreased frequency of the outcome of interest if the prevalence of exposure differs between cases and controls [1, 2].

N. A. Ghafar (✉)
School of Health Sciences, Universiti Sains Malaysia (Health Campus),
Kubang Kerian, Kelantan, Malaysia
e-mail: norainiag@usm.my

2   Method of Case–Control Studies

Case–control study is by default a retrospective in nature. However, a special type of case–control study, known as nested case–control study (discussed later in this chapter), is a hybrid design (partly prospective and partly retrospective). In traditional case–control study, cases and controls are selected at the starting point of the study – cases having the disease of interest, and control not having the diseases of interest. Then, researchers go back to collect information about the exposure status for both cases and controls. The association between the exposure and outcome is calculated after getting the exposure data of the participants [3, 4]. The method of case–control study is shown in Fig. 3.1 [5].

Before a case–control study is carefully planned, as with any other study type, the precise hypothesis being investigated must be articulated. Failure to do so may result in poor design and issues with result interpretation. Case–control studies enable the assessment of a variety of exposures that may be connected to a particular disease [3].

Fig. 3.1 Case–control study design [5]. Used with permission of John Wiley & Sons, Inc. from Chapter 17: Investigating the Types of Epidemiologic Studies, Mitra AK, first edition, 2023; permission conveyed through Copyright Clearance Center, Inc
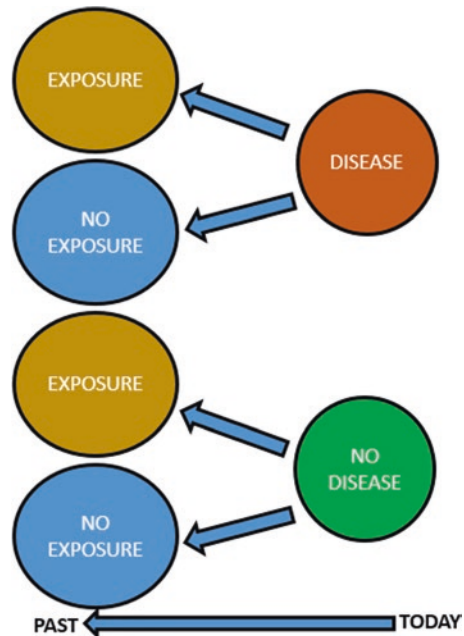
2.1 Selection of Cases

In a case control study, cases must be defined explicitly. Researchers must pay attention to define the cases as precise (and not ambiguous) as possible. For practical purposes, it is frequently useful to divide the diagnostic continuum into "cases" and "non-cases." Several issues should be considered in determining the cut off for such a division.

For a statistical point of view, the typical practice is to define "normal" as being within two standard deviations of the mean value. This is just an approximate guide to dividing cases from controls.

For the clinical option, certain clinical signs, such as high systolic blood pressure or low glucose tolerance, may be asymptomatic but have a negative prognosis. This may create bias in defining cases and control solely based on clinical parameters. Let us take another example: although normal blood pressure for most adults is defined as a systolic pressure of less than 120 and a diastolic pressure of less than 80, sometimes a male aged 50 or above may have a systolic blood pressure of ≈140 mm Hg but clinically normal in the absence of symptoms.

Nonetheless, an objective case definition must be used to differentiate cases from control. An investigator should have a distinct goal and purposes for the study before defining cases. Regardless of the method of selection of cases, the case definition should be as explicit and unambiguous as feasible [6].

2.2 Selection of Controls

The selection of controls is the next crucial aspect of designing a case–control study. It is important to select controls that are broadly comparable to the cases. The chosen control group needs to be at least similar in the likelihood of developing the result.

It is typical to employ two sources of controls, population control and hospital controls. The advantage of choosing controls from the general population is that their exposures probably represent individuals who are likely to develop cases. Population control is the most desirable method. Non-cases are sampled from the source population giving rise to cases. Another method called neighborhood controls or relative controls is recommended, provided they do not share the exposure of interest, such as smoking in the investigation of cancer.

The use of hospital controls is generally discouraged because of several issues. First, hospital controls may have diseases resulting from the exposure of interest. For example, smoking is an important risk factor for cancer. If we select cases having cancer, the controls may have a disease that is also related to smoking (such as asthma, COPD or heart disease). Secondly, hospital controls may not be representative of the exposure prevalence of the source population of cases. For example, smokers with some illnesses may be more hospitalized than the general population

who are not hospitalized. However, hospital controls are a vital source of controls. These controls are simple to recruit and are more likely to have medical records of comparable quality.

One of the other issues is bias in case–control studies. Cases are more motivated to recall facts of their past events than controls who have no particular interest in the research subject because they are eager to learn what caused their disease [4]. Selecting hospital controls with conditions believed to lead to comparable memory errors may mitigate some of the issues caused by this type of information bias [1]. More information about bias is discussed in Chap. 11.

2.3 Issues in Selecting Controls

Before looking at the issues in control selection, we have to first acknowledge the importance of comparability between cases and control. In case–control study, controls must be comparable with cases. Issues arise when controls were not comparable to cases. Sometimes, selection bias occurs in selecting the controls. The consequence of this is inaccurate results of the analysis.

Four strategies could be used to overcome the problem of selection bias. One of the four strategies may be employed to allow the controls to represent the same population as the cases.

1. A convenience sample – This is one of the most common methods of selecting samples. A convenience sample is drawn like that of the cases, such as by enrolling in the same outpatient department. While undoubtedly convenient, this could weaken the study's external validity.
2. Matching – A matched or unmatched random sample from the unaffected population may serve as the controls. Once more, there are issues with adjusting for unidentified influences, but if the controls are too similar, they might not be representative of the broader population. "Over matching" could lead to an underestimation of the real difference. The benefit of matching is that it enables any given impact to be statistically significant with a smaller sample size.
3. Using a minimum of two control groups – More than one control increases the statistical power of the study. This issue is further discussed later in this chapter. With having more controls, the conclusion is stronger if the study shows a substantial difference between the patients with the desired outcome and those without it, even when the latter group has been sampled in a variety of ways (for example, outpatients, inpatients, and general practice patients).
4. Both patients and controls are drawn from a population-based sample – a random sample of all patients with a particular ailment can be drawn from specified registers. The control group can then be created by choosing randomly selected individuals with similar age and sex distributions from the same population as the area covered by the disease registration.

Meanwhile, researchers can avoid the problem of observation and recall bias by utilizing the blinding technique. A double blinding method is one in which neither the subject nor the observer is aware of their status as a case or control subject. They are also not aware of the study's main objectives. However, blinding subjects to their case or control status is usually impractical as the subjects already know that they have a disease or illness [3, 7]. Instead, only partial blinding can be done. Asking fictitious questions typically allows one to blind the subjects and observers to the study hypothesis.

2.4   How Many Controls Suitable for Each Case

Finding a reliable source of cases is typically not too difficult, but choosing controls is more challenging. Controls should ideally meet two criteria. Their exposure to risk factors and confounders should, within the bounds of any matching criteria, be typical of that in the population "at risk" of becoming cases, or those who do not have the disease under research but would be included in the study as cases if they did. The exposures of controls should also be quantifiable with accuracy comparable to that of the cases. It frequently becomes impossible to accomplish both of these goals [4].

When both cases and controls are freely available, it is most efficient to select at least an equal number of each. However, the rarity of the disease being investigated frequently limits the number of instances that may be studied. In this situation, statistical confidence can be strengthened by incorporating multiple controls per cases. There is, however, a law of diminishing returns. Researchers designing case–control studies are typically recommended to include no more than four controls per case because adding additional controls does not add much statistical power by increasing this ratio. Among the factors to be considered when deciding the number of control for a matched case–control are [1] the desired type I error rate, [2] the minimum odds ratio to be detected as statistically significant, [3] the estimated number of cases, [4] the control-to-case ratio in the population, (5) the estimated prevalence of exposure in the control group, and [6] an estimate of the correlation coefficient for exposure between cases and their matched controls [8].

2.5   Advantages and Disadvantages of Case–Control Studies

Case–control studies are typically rapid, inexpensive, and simple to conduct. Samples of cases and controls are frequently taken from sources like an existing database of patient health records. Additionally, case–control studies are particularly well suited for researching the risk factors linked to uncommon diseases or conditions. In contrast, if the illness or condition is uncommon, an observational design, such as a prospective cohort study, would not be appropriate because it is

unlikely that many participants will experience the illness or condition of interest. Contrary to cohort research, case–control studies are less likely to have loss to follow-up. Before doing more extensive and expensive studies (such as cohort study), case–control studies are sometimes conducted as preliminary research to determine any potential correlations. Case–control studies have the additional drawback of being unsuitable for situations when exposure to any of the risk variables is uncommon because very few, if any, of the cases or controls are likely to have been exposed to them [9].

The biases and interpretation issues that affect all observational epidemiological studies also apply to case–control studies. Confounding, bias in selection or sampling, measurement error, and missing data are a few of these issues. Selection bias is a severe form of bias resulting from missing data in which respondents from the source population who are not included in the study have no observational data at all.

2.6   *Nested Case–Control Study*

The nested case-cohort study is an observational design that incorporates the case–control approach within an established cohort. The design overcomes some of the disadvantages associated with case–control studies while incorporating some of the advantages of a cohort study. For example, in a nested case–control study, the exposure factors such as blood samples for parameters that may determine a disease are already preserved. In this design, researchers start with a suitable cohort that contains a sufficient number of cases (to have sufficient statistical power to address the research topic). The researchers then decide on a random basis a representative sampling of the individuals who have no outcome or the condition being studied (the controls); they pick two or three controls to match with a case. This is done to improve the power of the study [10]. The process of case and control selection is done prospectively in a defined period of time (for example, 5 years). Once cases and controls have been selected, then researchers go back to analyze the already collected samples for laboratory tests.

Using nested case–control studies is a highly effective method for determining the causes of variability in cancer incidence rates within a community. Since a disease-free group of samples within a cohort is selected to begin with, there is less chance of the selection biases that can occur in a conventional case–control study (described earlier), which is solely retrospective. Due to the fact that data gathered as part of a cohort study are collected before the onset of sickness, information bias is less likely to occur. By limiting data extraction and coding to the nested case–control sample, substantial cost savings can frequently be achieved [11].

Other source data may also be collected "retrospectively" on the sampled participants, although the risk of information bias must be taken into account. In comparison to the analysis of the laboratory data for the entire cohort which is typical for a cohort study, a nested case–control sampling does not require such analysis of data of the entire cohort. The increase in efficiency is determined by the number of

controls per case. In many instances, such as examining the association between a disease and a rare exposure, evaluating the impact of confounding, or determining the variance in relative risk with a putative effect modifier, a specified level of efficiency may require a large number of controls [12]. However, a cost-effective analysis is required to select the number of controls per cases [13].

3 Example of Case–Control Studies

The following examples of case–control studies have been taken from published sources [14–17]:

Example 1 Tan et al. (2018) investigated breast cancer risk variables in Malaysian women [14]. Participants in the study are drawn from two hospitals in Selangor, Malaysia: the University Malaya Medical Centre (UMMC), a public hospital, and the Subang Jaya Medical Centre (SJMC), a private hospital. All patients with clinically diagnosed breast cancer were eligible to be included as cases. Cases from UMMC have been recruited since October 2002, while SJMC cases have been recruited from September 2012. Healthy women aged 40 to 74 with no history of breast cancer were recruited for the Malaysian Mammography Study (MyMammo) at UMMC and SJMC. MyMammo at SJMC is a subsidized opportunistic mammography screening initiative that began in 2011. At UMMC, MyMammo began recruitment in 2014 from patients attending normal opportunistic screening. All participants in the study were interviewed by trained interviewers at the hospitals. The participants filled out a questionnaire that asked about their demographics, personal and family history of cancer, history of breast surgery, menstrual and reproductive history, use of oral contraceptives and hormone replacement therapy (HRT), breast cancer diagnosis (cases only), and history of and motivation to attend mammography screening (controls) only. Participants supplied a blood sample, which was processed and stored. After controlling for demographics and other risk variables, participants who had breast surgery to remove cysts and lumps were 2.3 times (95% CI, 1.82 to 2.83) more likely to get breast cancer than those who had never had breast surgery. After controlling for demographics and other risk factors, a first-degree family history of breast cancer was related with a 19% increased risk of breast cancer. After controlling for demographic and other risk variables, "post-menopausal women had a 52% increased risk of breast cancer" [14]. Furthermore, the researchers determined that "breastfeeding, soy consumption, and physical exercise are modifiable risk factors for breast cancer" [14].

Example 2 This example of case–control study was conducted by Ganesh and colleagues in 2011 [15]. Only male patients were included in the study. Patients were interviewed at Tata Memorial Hospital's (TMH) outpatient department in Mumbai, India. The data was collected using a predesigned questionnaire that was pre-tested at the hospital. The questionnaire included demographic variables (age, gender, reli-

gion, etc.), lifestyle (habits such as smoking, chewing, drinking alcohol, etc.), dietary habits, and occupational exposure. Patients from all over India come to the hospital because it is a comprehensive cancer clinic for diagnosis and treatment. In general, 30–40% of overall registrations are diagnosed as cancer-free each year. These cancer-free patients were used as "controls" after their medical history and diagnoses were examined. The patients were lung cancer cases that had been microscopically proven. Controls were defined as those that were diagnosed by microscope as "free of cancer" and not having any respiratory tract diseases and therefore diagnosed as "no indication of disease." Major risk factors that showed a dose–response correlation with lung cancer included: cigarette smoking (OR = 5.2), bidi smoking (OR = 8.3), and alcohol drinking (OR = 1.8). Only red meat consumption indicated a 2.2-fold increased risk among the dietary categories studied. Milk consumption was associated with a 60% reduction in risk, whereas coffee was associated with a twofold increase in risk of lung cancer. Furthermore, pesticide use was linked to a 2.5-fold increased risk of lung cancer [15].

Example 3 Xi and colleagues (2020) investigated the relationship between maternal lifestyle and the risk of low birth weight in both preterm and term babies [16]. This case–control study was carried out in 14 Chinese hospitals in Jiangmen, Guangdong Province. A stratified sampling strategy was used based on geography. The number of deliveries was used to make a purposeful selection. Hospitals were picked to ensure that each region had at least two hospitals. From August 2015 to May 2016, the patients and controls in this study were recruited from the same hospitals. The researchers found that women who delivered preterm and were physically active (1–3 times per week and 4 times per week, respectively) had a lower risk of having low birth weight babies (aOR = 0.584, 95% CI = 0.394 to 0.867 and aOR = 0.516, 95% CI = 0.355 to 0.752). Pregnant women who did not gain enough gestational weight had a higher risk of having low birth weight kids (aOR = 2.272, 95% CI = 1.626 to 3.176). Women who were exposed to passive smoking had a higher risk of having low birth weight babies (aOR = 1.404, 95% CI = 1.057 to 1.864). For term deliveries, both insufficient gestational weight increase and excessive gestational weight gain were significantly linked with low birth weight (aOR = 1.484, 95% CI = 1.103 to 1.998 and aOR = 0.369, 95% CI = 0.236 to 0.577, respectively). Furthermore, "parity, a history of low birth weight, prenatal treatment, and gestational hypertension were all related with a higher risk of low birth weight" [16].

Example 4 Shimeles and colleagues (2019) conducted a case–control study to ascertain the risk factors of tuberculosis (TB) in Ethiopia [17]. In the study, the cases were newly detected bacteriologically confirmed pulmonary TB patients aged >15 years, enrolled for treatment in the selected health centers in Addis Ababa. Controls were age- and sex-matched attendees who presented in the same health centers for non-TB health problems. The data collection took place by including all newly registered TB patients until the required sample size was reached. In the study, it was revealed that patients who lived in houses with no window or one win-

dow (suggesting poor ventilation) were almost two times more likely to develop tuberculosis compared to people whose houses had multiple windows (aOR = 1.81; 95% CI =1.06 to 3.07). Besides, previous history of hospital admission was found to pose risk more than three times (aOR = 3.39; 95% CI = 1.64 to 7.03). Having a household member who had TB was shown to increase risk of developing TB by threefold (aOR = 3.00; 95% CI = 1.60 to 5.62). The study also showed that illiterate TB patients were found to be more than twice more likely to develop TB compared to subjects who can at least read and write (aOR, 95% CI = 2.15, 1.05 to 4.40). Patients with household income of less than 1000 Birrs (1 Birr = 0.018 US dollar) per month were more than two times more likely to develop TB compared to those who had higher income (aOR = 2.2; 95% CI = 1.28 to 3.78). Tobacco use was found as a fourfold risk factor for getting TB (aOR = 4.43; 95% CI = 2.10 to 9.3). BCG vaccination, on the other hand, was found to be protective against TB, lowering the risk by one-third (aOR = 0.34; 95% CI = 0.22 to 0.54) [17].

4 Calculation of Odds Ratio

Case–control studies produce the odds ratio as a measure of the degree of the association between an exposure and the outcome. It is the measure of association that compares the probabilities of disease or an occurrence among those who have been exposed to those who have not (Table 3.1). Its purpose is to establish the association between exposure and outcome [18].

Here, A = number of exposed subjects and they have the disease; B = number of exposed subjects but they do not have the disease; C = number of unexposed subjects and they have the disease; D = number of unexposed subjects and they do not have the disease.

$$\text{Odds ratio}(\text{OR}) = \frac{AD}{BC}$$

OR is a quantitative representation of the strength of association between a cause and an effect when both variables are presented as categorical variables. As a general rule, the greater the OR, the greater the effect on the outcome. OR is interpreted as follows:

Table 3.1 Calculation of odds ratio for case–control study

	Disease status		
Exposure status	Case	No disease	Total
Yes	A	B	$A + B$
No	C	D	$C + D$
Total	$A + C$	$B + D$	$A + B + C + D$

Table 3.2 Food poisoning and contaminated salad

Category	Cases (food poisoning)	Controls (without food poisoning)	Total
Exposed (ate contaminated salad)	15 (a)	9 (b)	24
Unexposed (did not eat contaminated salad)	8 (c)	32 (d)	40
Total	23	41	64

- OR of 1: There is no difference between the groups; i.e., there would be no association between the exposure (pizza) and the outcome (being ill).
- OR of >1: Suggests that the odds of exposure are positively associated with the adverse outcome compared to the odds of not being exposed.
- OR of <1 Suggests that the odds of exposure are negatively associated with the adverse outcomes compared to the odds of not being exposed [19].

OR is further illustrated by using real-life data in Table 3.2. Based on the table, those who ate the contaminated salad (exposure) were 6.67 times more likely (OR = 6.67) to get food poisoning (outcome), compared to those who did not eat the salad.

$$OR = \frac{ad}{bc} = \frac{15 \, x \, 32}{9 \, x \, 8} = 6.67$$

4.1 Calculation of 95% Confidence Intervals for OR

Each odds ratio should have a confidence interval (CI) calculated for it. A CI that includes 1.0 indicates that there is no statistically significant correlation between the exposure and the outcome, and that the correlation might have been obtained by chance alone. Without a confidence interval, an odds ratio is not very meaningful [20]. The 95% confidence interval (CI) is used to estimate the precision of the OR [21].

Based on Table 3.2, CI for the OR could be computed by using the following formula:

$$Upper \, 95\% CI = e^{\ln(OR) + 1.96\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}}$$

$$Lower \, 95\% CI = e^{\ln(OR) - 1.96\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}}$$

You can also use the following link for a quick calculator: https://www.medcalc.org/calc/odds_ratio.php.

Using the formula (or the quick calculator link), the upper limit of 95% CI = 20.70, and the lower limit of 95% CI = 2.15. In other words, the OR in the population from which the sample was drawn varied from 2.15 to 20.70.

5 Further Practice

1. Following are the advantages of case control study, except.

   (a) Multiple exposures or risk factors can be examined.
   (b) Rates of disease in exposed and unexposed could be determined.
   (c) Relatively quick to conduct.
   (d) Can use existing records.

2. Choose the advantage of matching in case–control studies:

   (a) Decision to match confounding variables is decided at the outset of the study.
   (b) Requires a matched analysis.
   (c) Eliminate influence of measurable confounders.
   (d) Matched variables cannot be examined in the study.

3. Following are techniques that could be used to ensure that the controls to represent the same population as the cases, except.

   (a) Using a convenience sample.
   (b) Blinding.
   (c) Using two or more control groups.
   (d) Using a population-based sample for both cases and controls.

4. Following is the purpose of matching in case control study, except.

   (a) To improve study efficiency by improving precision.
   (b) To enable control in the analysis of unquantifiable factors.
   (c) To eliminate sampling bias.
   (d) To make outcome groups comparable on the matching variable.

5. Following are the factors to be considered when deciding the number of control for a matched case control, except.

   (a) The desired type I error rate.
   (b) The maximum odds ratio to be detected as statistically significant.
   (c) The estimated number of cases.
   (d) The control-to-case ratio in the population.

6. Choose the reason why it is not advisable to use hospital controls.

   (a) Hospital controls may have diseases resulting from the exposure of interest.

(b) Hospital controls may be representative of the exposure prevalence of the source population of cases.

(c) The exposure of hospital controls may not be comparable with that of cases.

(d) Findings from studies using hospital controls tend to overestimate risk because of differential recall.

7. If a control group member had the ailment under investigation, he or she would have been recognized as a prospective case for the study. This is the fundamental notion that must be followed while picking an appropriate control group.

   True/False.

8. Adopting hospital-based controls in a case–control study has the advantage of minimizing selection bias.

   True/False.

9. A risk ratio or an odds ratio can be calculated in a case–control study.

   True/False.

10. If you use more than four controls for each case, the power of the study is much increased.

   True/False.

Answer Keys
1. (d)
2. (c)
3. (b)
4. (c)
5. (b)
6. (a)
7. False
8. True
9. False
10. False.

References

1. Lewallen S, Courtright P. Epidemiology in practice: case-control studies. Community Eye Health. 1998;11(28):57–8. PMID: 17492047; PMCID: PMC1706071
2. dos Santos Silva I. 1999. Chapter 9: case control studies in cancer epidemiology: principles and methods ISBN-13 978–92–832-0405-3.
3. Breslow NE. Statistics in epidemiology: the case-control study. J Am Stat Assoc. 1996;91(433):14–28. https://doi.org/10.1080/01621459.1996.10476660. PMID: 12155399
4. Setia MS. Methodology series module 2: case control studies. Indian J Dermatol. 2016;61:146–51.

5. Mitra AK. Investigating the types of epidemiologic studies. In: Epidemiology for dummies. 1st ed. Hoboken, New Jersey, United States: Wiley; 2023.

6. Critchley J. Epidemiology for the uninitiated, 5th ed. J Epidemiol Community Health. 2004;58

7. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies. I Principles Am J Epidemiol. 1992;135(9):1019–28. https://doi.org/10.1093/oxford-journals.aje.a116396. PMID: 1595688

8. Hennessy S, Bilker WB, Berlin JA, Strom BL. Factors influencing the optimal control-to-case ratio in matched case-control studies. Am J Epidemiol. 1999;149(2):195–7. https://doi.org/10.1093/oxfordjournals.aje.a009786. Erratum in: Am J Epidemiol 1999 Mar 1;149(5):489. PMID: 9921965

9. Sedgwick P. Case-control studies: advantages and disadvantages. BMJ. 2014;3(348):f7707. https://doi.org/10.1136/bmj.f7707. PMID: 31419845

10. Sedgwick P. Nested case-control studies: advantages and disadvantages. BMJ (online). 2014;348:g1532. https://doi.org/10.1136/bmj.g1532.

11. Sedgwick P. Nested case-control studies. BMJ. 2010;(340):c2582. https://doi.org/10.1136/bmj.c2582. PMID: 20484347

12. Ernster. Nested case control studies. Prev Med. 1994;23:587–90.

13. Langholz B, Clayton D. Sampling strategies in nested case-control studies. Environ Health Perspect. 1994;102 Suppl 8(Suppl 8):47–51. https://doi.org/10.1289/ehp.94102s847. PMID: 7851330; PMCID: PMC1566552

14. Tan MM, Ho WK, Yoon SY, Mariapun S, Hasan SN, Lee DS, Hassan T, Lee SY, Phuah SY, Sivanandan K, Ng PP, Rajaram N, Jaganathan M, Jamaris S, Islam T, Rahmat K, Fadzli F, Vijayananthan A, Rajadurai P, See MH, Thong MK, Mohd Taib NA, Yip CH, Teo SH. A case-control study of breast cancer risk factors in 7,663 women in Malaysia. PLoS One. 2018;13(9):e0203469. https://doi.org/10.1371/journal.pone.0203469. PMID: 30216346; PMCID: PMC6138391

15. Ganesh B, Sushama S, Monika S, Suvarna P. A case-control study of risk factors for lung cancer in Mumbai. India Asian Pac J Cancer Prev. 2011;12(2):357–62. PMID: 21545194

16. Xi C, Luo M, Wang T, Wang Y, Wang S, Guo L, Lu C. Association between maternal lifestyle factors and low birth weight in preterm and term births: a case-control study. Reprod Health. 2020;17(1):93. https://doi.org/10.1186/s12978-020-00932-9.

17. Shimeles E, Enquselassie F, Aseffa A, Tilahun M, Mekonen A, Wondimagegn G, Hailu T. Risk factors for tuberculosis: a case-control study in Addis Ababa, Ethiopia. PLoS One. 2019;14(4):e0214235. https://doi.org/10.1371/journal.pone.0214235. PMID: 30939169; PMCID: PMC6445425

18. Kalra A. The odds ratio: principles and applications. J Pract Cardiovasc Sci. 2016;2:49–51.

19. Szumilas M. Explaining odds ratios. J Can Acad Child Adolesc Psychiatry. 2010;19(3):227–9. Erratum in: J Can Acad Child Adolesc Psychiatry. 2015 Winter;24(1):58. PMID: 20842279; PMCID: PMC2938757

20. Mann CJ. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. Emerg Med J. 2003;20:54–60.

21. Song JW, Chung KC. Observational studies: cohort and case-control studies. Plast Reconstr Surg. 2010;126(6):2234–42. https://doi.org/10.1097/PRS.0b013e3181f44abc. PMID: 20697313; PMCID: PMC2998589