

# General concepts in biostatistics and clinical epidemiology: observational studies with case-control design

Diego Martínez<sup>a</sup>, Cristian Papuzinski<sup>a,b</sup>, Jana Stojanova<sup>a,b</sup>, Marcelo Arancibia<sup>a,b,\*</sup>

<sup>a</sup> Cátedra de Metodología de la Investigación Científica, Escuela de Medicina, Universidad de Valparaíso, Viña del Mar, Chile

<sup>b</sup> Centro Interdisciplinario de Estudios en Salud (CIESAL), Universidad de Valparaíso, Valparaíso, Chile

\*Corresponding author marcelo.arancibiame@uv.cl

**Citation** Martínez D, Papuzinski C, Stojanova J, Arancibia M. General concepts in biostatistics and clinical epidemiology: observational studies with case-control design. *Medwave* 2019;19(10):e7716

**Doi** 10.5867/medwave.2019.10.7716

**Submission date** 10/9/2019

**Acceptance date** 25/10/2019

**Publication date** 11/11/2019

**Origin** This article is one of several “Methodological notes” prepared by the course on Research Methodology of the School of Medicine of the University of Valparaíso

**Type of review** Externally peer-reviewed by three reviewers, double-blind.

**Key words** observational study, case-control studies, bias, epidemiology, biostatistics

## Abstract

Case-control studies have been essential to the field of epidemiology and in public health research. In this design, data analysis is carried out from the outcome to the exposure, that is, retrospectively, as the association between exposure and outcome is studied between people who present a condition (cases) and those who do not (controls). They are thus very useful for studying infrequent conditions, or for those that involve a long latency period. There are different case selection methodologies, but the central aspect is the selection of controls. Data collection can be retrospective (obtained from clinical records) or prospective (applying data collection instruments to participants). Depending on the objective of the study, different types of case-control studies are available; however, all present a particular vulnerability to information bias and confounding, which can be controlled at the level of design and in the statistical analysis. This review addresses general theoretical concepts concerning case-control studies, including their historical development, methods for selecting participants, types of case-control studies, association measures, potential biases, as well as their advantages and disadvantages. Finally, concepts about the relevance on this study design are discussed, with a view to aid comprehension for undergraduate and graduate students of the health sciences. This is the third of a methodological series of articles on general concepts in biostatistics and clinical epidemiology developed by the Chair of Scientific Research Methodology at the

School of Medicine, University of Valparaíso, Chile.

## Key ideas

- Case-control studies analyze associations between an exposure factor and an outcome in people in whom the condition is present (cases) or absent (controls).
- This design is useful for studying conditions that are infrequent, or those that require a long latency to occur. However, they are vulnerable to information bias and confounding.

## Introduction

Elements of the case-control design have been evident since the nineteenth century. Perhaps the most well-known example is that of the cholera outbreaks investigated by John Snow and Reverend

Henry Whitehead, ultimately leading to the discovery that the Broad Street water pump was the cause<sup>1,2</sup>. Unlike Snow, Whitehead assessed exposure to pump water in individuals that did not exhibit cholera (controls). Through a thorough and systematic survey,

which included visiting individuals up to five times, Whitehead collected basic but relevant information regarding water consumption among Broad Street residents, concluding that using water from a specific pump associated with cholera, a finding that resulted in a decrease from 127 deaths on September 2, to 30 on September 8, in 1854<sup>3</sup>.

However, the modern conception of the case-control design is attributed to Janet Lane-Clayton for her work on risk factors associated with breast cancer (1926)<sup>4</sup>. In 1939, another case-control study led by Franz Müller<sup>5</sup>, member of the Nazi party, linked the consumption of cigarettes with lung cancer, consistent with Hitler's position against smoking; indeed, his government promoted propaganda campaigns against tobacco consumption in light of recently available evidence. Müller sent a questionnaire to relatives of lung cancer victims, inquiring about consumption habits, including form, frequency, and type of tobacco used, corroborating a strong association between tobacco consumption and the disease<sup>5,6</sup>. Subsequently, and parallel to the course of World War II, there was a halt in the development of this methodological design until four case-control studies were published in 1950. They all analyzed the relationship between smoking and lung cancer, validating the use of this design to determine the etiology of diseases. One of these was led by Richard Doll and Austin Bradford Hill<sup>7,8</sup>, who believed that increases in lung cancer rates in England and Wales could not fully be explained by improvements in diagnostic tests -as was argued at the time- but rather environmental factors including smoking and air pollution<sup>7</sup>.

Decades later, in 1987, a study of risk factors associated with the transmission of Acquired Immunodeficiency Syndrome, such as promiscuity and the use of intravenous drugs<sup>9,10</sup>, enabled the implementation of measures that reduced transmission, even before the virus had been identified<sup>10</sup>.

Thus, epidemiology shifted from determining causes to determining risk factors<sup>1</sup>; Snow was not interested in determining the causal agent but rather ways cholera was transmitted<sup>3</sup>. In this way, observational designs such as case-control and cohort studies are available to study etiology and prognostic factors (protective factors and risk factors)<sup>11</sup>. In this article, we will focus on the former, while cohort studies will be the subject of the next article in this series.

This review is the third of a methodological series comprising six narrative reviews that cover general topics in biostatistics and clinical epidemiology. The series is based on content from publications available from major databases of the scientific literature, as well as specialized reference texts. The series is oriented toward undergraduate and graduate students and is developed by the Chair of Scientific Research Methodology at the School of Medicine, University of Valparaíso, Chile. Therefore, the purpose of this manuscript is to address the main theoretical and practical concepts of case-control studies.

## Preliminary concepts

Case-control studies constitute an observational, analytical and longitudinal design: the researcher does not assign exposures, the design permits hypothesis testing, and there is a period between exposures

and outcomes. Some authors purport that causal relationships could be demonstrated through a case-control design<sup>12</sup>; however, this is controversial. To execute a case-control study, a group of participants similar in baseline characteristics are recruited that either present an outcome of interest (cases) or do not present it (controls). In both cases and controls, variables that represent risk factors are measured and compared between the two. Thus, a fundamental characteristic of a case-control study is that the subjects are selected according to an outcome; this is an advantage given it is not necessary to wait a prolonged period for the phenomenon under study to occur. It can also present a disadvantage as in most cases data will have to be collected retrospectively and quality will depend on an adequate record or participants' memory<sup>13</sup>.

### Selection of cases

The selection of cases must be rigorous, privileging incident cases (cases that have been recently diagnosed) over prevalent cases (all available cases, including those diagnosed years prior). Incident cases are likely more similar in how they were diagnosed, and more consistent with the present diagnostic criteria.

It is thus necessary to have a clear definition of the outcome, for example, current and international diagnostic criteria, laboratory tests, imaging studies, among others. This is supported by clearly stated eligibility criteria, such as enrollment site and age range<sup>14,15</sup>. Potential sources for cases include hospitals, communities or population registries, or patient groups, such as Alcoholics Anonymous or support groups such as those for specific genetic diseases. Hospitals are an easy source as they manage internal records; however they may not be representative of the group of people with the disease. On the other hand, population cases are more challenging to locate in the absence of registries but present the advantage of being more representative<sup>16</sup>.

### Selection of controls

Selection of controls is a crucial aspect of case-control studies as an investigation's internal validity depends on it. Controls represent the baseline frequency of exposures in individuals free of the outcome under study. It is important not to limit the selection of controls to healthy subjects; the fundamental aspect is absence of the disease outcome under study, independent of the presence or absence of risk factors of interest<sup>17</sup>.

Controls must be representative of the population from which cases were obtained, that is, they must be extracted from the same population base ("principle of the study base")<sup>16</sup> in order to present the same exposure risk as cases and avoid selection bias. Selection by random sampling is the best means to ensure controls have the same theoretical probability of exposure to risk factors as cases<sup>18</sup>. The number of controls for each case should not exceed three or four as increase in study power is minimal and disproportionate to the cost implied<sup>17,19</sup>. This corresponds to the "principle of efficiency", both statistical (achieving adequate power) and operational (optimizing the use of time, energy and research resources)<sup>16</sup>.

Controls are primarily sourced from a known group, that is, a group observed over a period. Nonetheless, the group from which cases are identified is often initially unknown, and the delimitation of the

group for selection of participants would, therefore, occur *a posteriori*<sup>20</sup>. Some strategies have been suggested for when the population base of cases is unknown, such as selecting controls that are neighbors of cases<sup>17</sup>. Likewise, it has been proposed that controls could be friends, thus share characteristics such as socioeconomic and educational level, or family members, thus share genetic and lifestyle characteristics. Selection of controls could also be made from other hospital patients, thus likely to come from a similar locality as controls, and present similar health-seeking behaviors versus controls sourced from the community<sup>20</sup>. However, hospital sourced controls might not share the same probability of exposures to risk factors as cases<sup>17</sup>.

Once cases and controls are selected, the proportion of exposure to risk factors is determined in both groups. These data can be sought in participants' clinical records or by applying questionnaires. In order not to incur biases in posterior analyses, the same thoroughness in sourcing data must be applied to cases and controls. Finally, to the extent that the difference in the proportion of participants exposed to a risk factor between the groups is greater, the greater the likelihood that there will be an association between the outcome and the exposure<sup>11</sup>.

### Measures of Association

Due to the nature of the case and control design, the measure of association is estimated in relation to an event that has already occurred, comparing the frequency of exposure between cases and controls, in addition to other estimators. Relative risk cannot be calculated due to the retrospective nature of the event, but rather an odds ratio is estimated with an associated confidence interval<sup>10</sup>.

This measure represents the ratio between the odds of exposure in the cases and controls, interpreted as how many times the odds of exposure are greater in cases compared controls: it is important to note that this does not represent a relative risk<sup>16</sup>. The odds ratio approximates relative risk when the disease or outcome is infrequent, for example, occurring at a prevalence no greater than 5% to 10% in both exposed and unexposed individuals. This is known as the "rare disease assumption"<sup>16</sup>.

The odds ratio has an interpretation similar -but not equal- to relative risk, taking values that range from zero to infinity. An odds ratio less than 1 indicates that the exposure behaves as a protective factor, while greater than 1 indicates a risk factor, that is, it increases the probability that the outcome will occur. Finally, if its value were equal to 1, it could be deduced that no association exists between exposure factor and outcome<sup>21</sup> (Example 1)<sup>1</sup>.

**Example 1.** In Whitehead's study of cholera at the Broad Street water pump, the odds ratio was calculated to be 19.6 for those who drank from the pump compared to those that did not. An odds ratio greater than 1 indicates a risk factor. It can be interpreted as follows: individuals who presented cholera (cases) had a 19.6 times greater *odds* of drinking water (retrospective interpretation) from the pump than those who did not present cholera (controls).

Through the cases-control design, the incidence or prevalence of a condition cannot be directly calculated. An exception would be pop-

ulation case-control studies, where it is recognized that the prevalence of exposure of the control group is representative of the entire population and the population incidence of the variable to be studied is known, permitting the estimation of the incidence. This estimate would be possible in case-control studies nested in a cohort and in case-cohort studies<sup>15</sup>: both of these design will be detailed below.

## Types of case-control studies

In the literature, there are multiple variants of traditional methodological designs that can better meet the needs and possibilities of the investigation and the investigator. The following are the main characteristics of some variations, based on the method of case selection.

### Case-control studies based on cases

This design corresponds to the traditional and most frequently performed type of case-control study. Existing (prevalent) or new (incident) cases are recruited, and a control group is formed from the same hypothetical cohort (hospital or population)<sup>16</sup>.

### Nested case-control studies

In this design, cases are selected among participants in a cohort study, that is, a prospective study where all the participants were initially free of the outcome of interest. Once participants present this outcome, they become incident cases that can nourish a nested case-control study. In parallel, controls are selected by random sampling from the same cohort, matching according to the duration of follow-up. This type of study is convenient as it offers better control of confounding factors since the cohort constitutes a homogeneous group defined in space and time. It also facilitates better quantification of the impact of time-dependent exposures, as the occurrence of the outcome is precisely known<sup>15,18</sup>.

### Cross-case, case-case or self-controlled studies (case-crossover studies)

In this recently developed methodological design, the exposure history of each patient is used as their own control (matched design), aiming to eliminate interpersonal differences that contribute to confounding<sup>22-24</sup>. This design is useful in the analysis of transient exposures, such as a period of poor sleep as a risk factor for car accidents. A "case period" is defined, which could be, for example, 48 hours before an accident. In this scenario, a "control period" might be between 72 and 49 hours prior. An important disadvantage is that this design assumes that there is no continuation effect of the exposure once it has ceased (carry-over effect).

### Case-cohort studies

This is a mixed design that involves characteristics of a case-control study and a cohort study; however, it is methodologically more similar to the latter<sup>25</sup>. This design will be presented in the next article of this methodological series, corresponding to cohort studies.

## Bias

In case-control studies, the characteristic with the greatest influence on biases is that the analysis starts from the outcome and not from the exposure, obtaining information mostly retrospectively. Biases that may occur during study planning require attention, such as undervaluing the economic cost of the study that may affect adequate completion<sup>26</sup>.

### Selection bias

Selection bias affects comparability between the groups studied due to a lack of similarity. Cases and controls will thus differ in baseline characteristics, whether these are measured or not, due to differential way of selecting them. It is thus necessary to ensure that cases and controls are similar in all important characteristics besides the outcome studied<sup>27</sup>. One example of selection bias is Berkson's paradox, also known as Berkson's bias, Berkson's fallacy, or admission rate bias<sup>26,27</sup>. For example, admission rates of cases that are exposed may differ in cases unexposed to the risk factor under study, affecting the risk estimate in cases (Example 2)<sup>28</sup>.

**Example 2.** Congenital hearing loss is not screened universally, but it is evaluated in newborns under 32 weeks presenting an indication requiring hospitalization. However, 50% of cases of congenital hearing loss are seen in newborns that do not present an indication for hospitalization, hence do not present risk factors that also associate with hospitalization. If a case-control study were conducted solely including hospital participants, cases of congenital hearing loss in term infants would be underrepresented.

Another type of selection bias is Neyman's bias<sup>26,27</sup>, also called prevalence-incidence bias. It occurs when a certain condition causes premature deaths preventing their inclusion in the case group, which may result in an association not being obtained due to the lack of inclusion in the analysis of participants who have already died. Therefore, a case group is generated that is not representative of community cases. Such is the case of diseases that are rapidly fatal, may exhibit subclinical presentations or are transient (Example 3).

**Example 3.** The relationship between arterial hypertension (risk factor) and stroke (outcome) is studied. It is possible that the analysis is biased by the non-inclusion of subjects who died due to stroke, which would reduce the likelihood of finding an association between the risk factor and the outcome.

### Information bias

Also called observation, classification or measurement bias. It appears when there is an incorrect determination of exposure or outcome<sup>27</sup>. In case-control studies, exposure information should be collected in the same way in both groups (known as the “principle of comparable accuracy”<sup>16</sup>) and ideally by trained people who do not know to which group participants respondents belong. Prior knowledge of case status may influence information gathering and may be known as interviewer bias<sup>14</sup>.

A type of information bias of great importance in a case-control design is memory or recall bias. Cases tend to search their memory for

factors that may have caused their disease, while controls are unlikely to have this motivation. Therefore, cases may remember exposures to the factors under study better than controls<sup>17</sup>. One solution that has been proposed is that controls with diseases similar to the one being studied ought to be selected. For example, if the case group has cancer A, the controls could have cancer B, so that similar recall tendencies occur between the groups. This procedure is valid so long as the exposure under study is known not to be related to the pathology present in the control group; otherwise it would contribute further bias. Other recommended methods include the use of memory aids such as photographs, calendars, newspapers, or any material that helps clarify recall of the exposure in the participants of both groups<sup>10</sup>.

### Confounding

This phenomenon has previously been addressed in two earlier articles of this methodological series<sup>29,30</sup>. Strategies to control confounding may be implemented at the level of methodological design (restriction and matching) and statistical analysis (stratified analysis, statistical regression and use of propensity scores).

The restriction procedure corresponds to the strict selection of subjects who present characteristics that investigators want to “neutralize” or who do not present them. Although this increases internal validity (by decreasing confounding), it also decreases external validity as the groups are less representative of the general population and results are less able to be extrapolated<sup>27</sup>.

Matching is another strategy to reduce confounding. It involves the selection of controls who share the characteristics to be neutralized present in cases, for example, similar socioeconomic level or age group<sup>31</sup>. For example, in a study that seeks to compare a group of women with and without multiple sclerosis, the first case is a carrier of the disease, is 40 years old and is of high socioeconomic status; the corresponding control would be a woman of the same characteristics but without the disease. This is detailed in Example 4, based on Lane-Clayton's<sup>4,6</sup> and Müller's<sup>5</sup> investigations.

**Example 4.** In Franz Müller's study on tobacco use and its association with lung cancer, a control group of the same size, sex and age range was recruited to match a group of 86 patients with the pathology. Similarly, in her study of women with breast cancer, Janet Lane-Clayton was careful to selected controls from hospital and non-hospital contexts in Glasgow and London. All participants were similar in age range and socioeconomic status. The study showed that fertility was 22% lower in women with breast cancer, the first evidence that suggested low fertility increased breast cancer risk. In both cases, the authors sought to “neutralize” the characteristics that could “confuse the association” because they were external to the causal pathway being investigated. In this way, they would be approaching the theoretical ideal that the only thing differentiating cases from controls is the presence of lung and breast cancer, respectively.

Applying matching carries a series of challenges: it prevents the analysis of variables that are used in matching<sup>27</sup> and can increase the time and cost of the study when appropriate controls are difficult to source<sup>32</sup>. Therefore, pairing should be carried out by variables that

represent legitimate potential confounding factors, since arbitrary variables will affect study efficiency and decrease validity of the comparison between cases and controls. This phenomenon is known as overmatching<sup>33</sup> and may affect the ability to detect differences between cases and controls that should be detected. It is therefore suggested that matching is not mandatory for case-control studies and is likely best applied in studies of limited number of cases or very rare exposures<sup>16</sup>.

Stratified analysis can be considered a *post hoc* form of restriction and involves the study of variables of interest stratified by levels of potential confounding variables. This is illustrated by Simpson's paradox<sup>16</sup>, a phenomenon in which an association measure between exposure and outcome, such as an odds ratio, is different when estimated across an entire group versus calculations within individual strata, such as age groups, sex, among others.

The Mantel-Haenszel method determines whether there is an association between an exposure and an outcome controlling the effect of one or more confounding factors. If the effect adjusted by the Mantel-Haenszel method differs significantly from the unadjusted or crude effect, it is presumed that the confounding factor is present<sup>14,27</sup>. Stratification may be limited by the sample size, and a single stratum may represent a very limited number of observations. A case of stratified analysis is presented in Example 5.

**Example 5.** Dysphagia is a recognized consequence of stroke, it is therefore decided that all stroke patients in a neurorehabilitation service will receive a nasogastric tube to avoid aspiration pneumonia. The odds ratio of dysphagia if stroke had occurred is 1.6 (95% confidence interval: 1.2 to 2.1). Stratified analysis was conducted: stratum A, patients in whom a nasogastric tube was left, and stratum B, patients in whom the nasogastric tube was withdrawn after otolaryngology evaluation. In stratum A, the odds ratio of having dysphagia if there was a stroke was 2.0 (95% confidence interval: 1.4 to 3.0); in stratum B, the odds ratio was 1.3 (95% confidence interval: 0.8 to 1.8). The crude odds ratio indicated a significant association between stroke and dysphagia (stroke was a risk factor). However, only in stratum A the odds ratio was statistically significant (the confidence interval did not include the number 1). The adjusted odds ratio, calculated by the Mantel-Haenszel method as a combination of odds ratio in A and B, resulted in 1.1 (95% confidence interval: 0.8 to 1.5). After controlling by nasogastric tube (confounding variable), it is concluded that stroke does not significantly associate dysphagia.

As has been covered in previous articles of this series<sup>29,30</sup>, confounding variables can also be addressed by multivariate regression. Its objective is to adjust a prediction model for a dependent variable by including multiple confounding variables<sup>34,35</sup>.

Finally, another strategy to address confounding in observational studies is the use of a propensity score<sup>36</sup>. Its purpose is to reduce confounding by indication (selection bias) and corresponds to the probability of treatment assignment conditional on baseline characteristics<sup>37</sup>. This score represents the probability of exposure estimated from a set of variables known to influence the probability of

exposure: the higher the score, the greater the probability of exposure. Subgroups of analysis can be stratified according to this score, or the score can be included as a covariate in multivariate statistical regression models<sup>35</sup>. It is important to consider that despite strategies to address confounding in the design and analysis of a study, some level of residual confounding may persist, especially in observational studies<sup>31</sup>.

## Advantages and disadvantages

Case-control studies are the best epidemiological design to investigate infrequent diseases, such as outbreaks, exemplified by the study of cholera associated with the Broad Street water pump. They are usually conducted quite quickly as outcomes have already occurred, leading to rapid results<sup>10</sup>. They are useful in pathologies involving a long latency period, which is prohibitive in other designs such as in a cohort study, as investigators will need to wait to observe the onset of the disease. Another positive aspect is that they allow the study of different risk factors simultaneously<sup>11</sup>.

A central challenge is a difficulty in determining the temporality of events, that is, if the cause preceded the effect, as would be expected. Another issue is the possibility of selecting controls in whom the pathology of interest is latent. This study design does not allow directly calculating risk since only the proportion of people that were exposed in case and control groups can be defined. Additionally, certain types of biases, such as recall bias, are particularly prominent<sup>10,14</sup>.

## Final considerations

Reading of reports of case-control studies should be done thoroughly as it may not be very intuitive to consider the measure of an association between a factor and an outcome starting from the latter, rather than the former. This design is therefore typically classified as retrospective, although some authors argue this is inaccurate given that data collection might be undertaken prospectively. It is therefore useful for authors to report whether the temporal classification, retrospective or prospective, is made according to the design or data collection strategy<sup>16</sup>.

Although it has been proposed that in selecting cases, diagnostic definitions must be clear and addressed in the eligibility criteria, it should be considered that multiple and very strict criteria will limit the external validity (or generalizability) of results. Nonetheless, the study must be planned on the premise that internal validity is a priority over external validity since the latter depends on the former<sup>16</sup>.

Regarding controls, some authors have forwarded the idea of using two groups of controls, selecting the group that presents the best characteristics after performing analyses<sup>17,38</sup>. The fundamental aspect is choosing controls, so they are similar to cases besides presenting the outcome of interest. In this line, some authors indicate that results of a case-control study should not be accepted until the reader assesses the rigor with which controls were selected<sup>14</sup>. Central to this is adequate reporting of the study, which can be achieved following guidelines such as STROBE (Strengthening the Reporting of Observational studies in Epidemiology) (<http://www.strobe-statement.org/>)<sup>39</sup>.

Case-control studies have strengths and have historically been a cornerstone in the study of major public health problems. However, their main weakness is that exposures occurred in the past, lending a particular sensitivity to bias. Confounding may be addressed by stratified analysis and the Mantel-Haenszel technique, but these have largely been replaced by multivariate statistical regression models<sup>40</sup>. Similarly, the use of matching has been diminished in favor of the use of statistical regression methods<sup>15,16</sup>. Nevertheless, the most advanced statistical analysis will not save a poorly designed study: controls must always be selected with maximum rigor.

## Notes

### Roles and contributions of authorship

CP, JS and MA are scholars in the Chair of Scientific Research Methodology, in which the development of this methodological series is circumscribed as a research activity of the teaching assistants of the course. All authors contributed to the planning and writing of the original manuscript.

DM, CP and MA contributed to the development of the Introduction, preliminary concepts, measures of association and types of case studies and controls. DM, MA and JS developed biases in case-control studies and also provided Examples 1, 3 and 4. DM, MA and CP provided Examples 2 and 5. DM and MA developed Figure 1.

### Funding

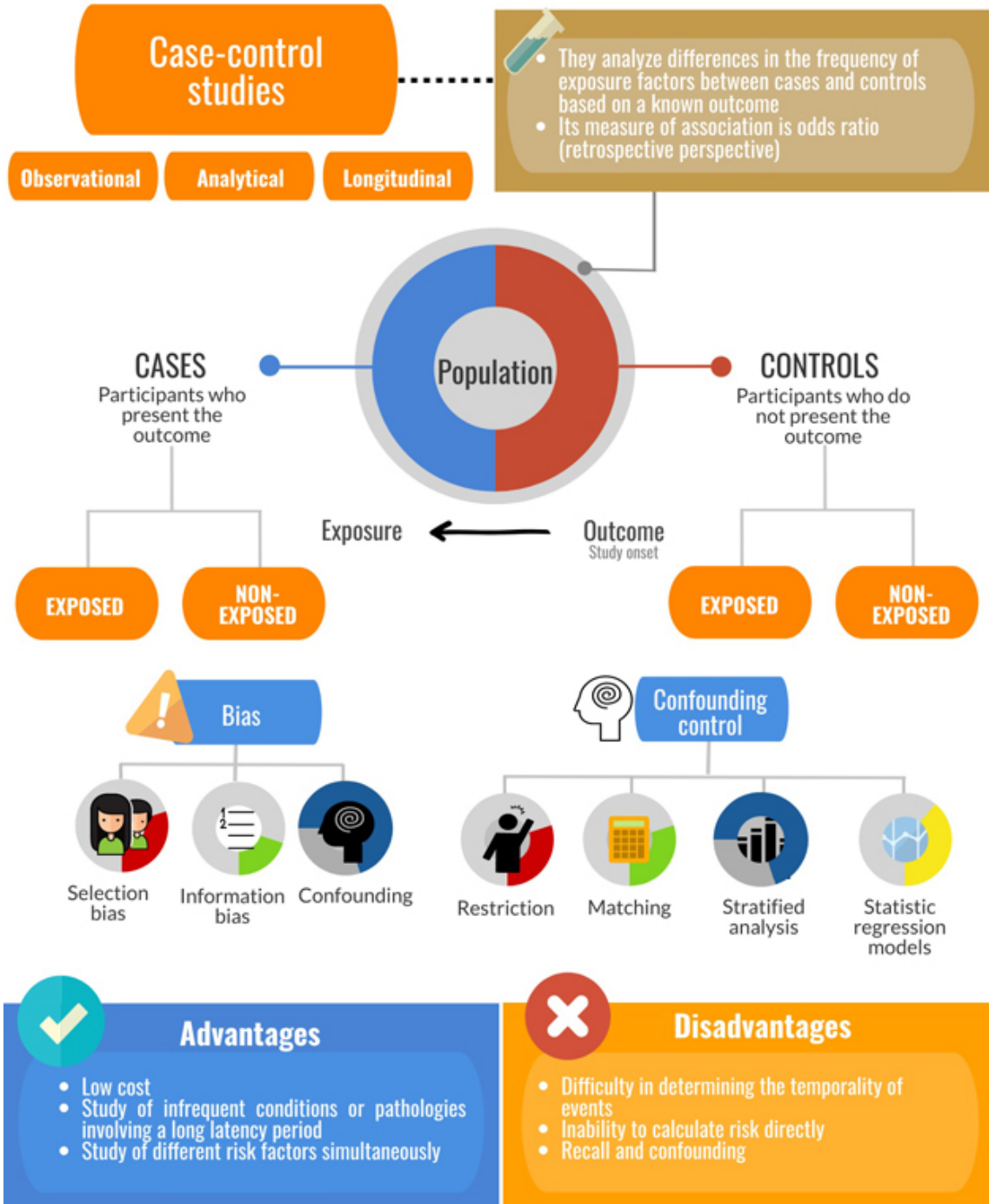
The authors declare that there were no external sources of funding.

### Competing interests

The authors have completed the ICMJE conflict of interest declaration form, and declare that they have not received funding for the completion of the report; have no financial relationships with organizations that might have an interest in the published article in the last three years; and have no other relationships or activities that could influence the published article. Forms can be requested by contacting the responsible author or the editorial board of the Journal.

# Case-control studies

## Concepts, bias, advantages and disadvantages



## References

1. Paneth N, Susser E, Susser M. Origins and early development of the case-control study: Part 1, Early evolution. *Soz Präventivmed.* 2002;47(5):282-8. | CrossRef | PubMed |
2. Shiode N, Shiode S, Rod-Thatcher E, Rana S, Vinten-Johansen P. The mortality rates and the space-time patterns of John Snow's cholera epidemic map. *Int J Health Geogr.* 2015 Jun 17;14:21. | CrossRef | PubMed |
3. Paneth N. Assessing the contributions of John Snow to epidemiology: 150 years after removal of the broad street pump handle. *Epidemiology.* 2004 Sep;15(5):514-6. | CrossRef | PubMed |
4. Lane-Clayton J. A further report on cancer of the breast. In: Reports on public health and medical subjects. London: Ministry of Health; 1926. | Link |
5. Müller F. Tabakmissbrauch und Lungencarcinom. *Ztschr Krebsforsch.* 1939;49:57-85.
6. Paneth N, Susser E, Susser M. Origins and early development of the case-control study: Part 2, The case-control study from Lane-Clayton to 1950. *Soz Präventivmed.* 2002;47(6):359-65. | PubMed |
7. Doll R, Hill AB. Smoking and carcinoma of the lung; preliminary report. *Br Med J.* 1950 Sep 30;2(4682):739-48. | CrossRef | PubMed |
8. Franceschi S. Sir Richard Doll. *Salud Publica Mex.* 2006;48 Suppl 1:S217-8. | CrossRef | PubMed |
9. Moss AR, Osmond D, Bacchetti P, Chermann JC, Barre-Sinoussi F, Carlson J. Risk factors for AIDS and HIV seropositivity in homosexual men. *Am J Epidemiol.* 1987 Jun;125(6):1035-47. | PubMed |
10. Papuzinski C, Martínez F. Case-control studies – the retrospective perspective. *Medwave.* 2014 Mar 27;14(2):e5925. | CrossRef | PubMed |
11. Araujo M. Etiology and prognostic clinical trials. *Medwave* 2011 Jun;11(06):e5052. | CrossRef |
12. Sackett D, Haynes R, Guyatt G, Tugwell P. *Clinical epidemiology: a basic science for clinical medicine.* 2nd ed. Boston: Little Brown and Company; 1991. | Link |
13. Araujo M. The temporality of clinical trials. *Medwave* 2011 May;11(05):e5020. | CrossRef |
14. Schulz KF, Grimes DA. Case-control studies: research in reverse. *Lancet.* 2002 Feb 2;359(9304):431-4. | PubMed |
15. Lazcano-Ponce E, Salazar-Martínez E, Hernández-Ávila M. Estudios epidemiológicos de casos y controles. Fundamento teórico, variantes y aplicaciones. *Salud Pública Mex.* 2001;43(2):135-50. | Link |
16. Gómez-Restrepo C. Estudios de casos y controles. In: Ruiz Á, Gómez-Restrepo C, editors. *Epidemiología clínica: investigación clínica aplicada.* 2nd ed. Bogotá: Editorial Médica Panamericana; 2015.
17. Grimes DA, Schulz KF. Compared to what? Finding controls for case-control studies. *Lancet.* 2005 Apr 16-22;365(9468):1429-33. | PubMed |
18. Langholz B, Richardson D. Are nested case-control studies biased? *Epidemiology.* 2009 May;20(3):321-9. | CrossRef | PubMed |
19. Kumar A, Dogra S, Kaur A, Modi M, Thakur A, Saluja S. Approach to sample size calculation in medical research. *Curr Med Res Pract.* 2014 Mar 1;4(2):87-92. | CrossRef |
20. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies. I. Principles. *Am J Epidemiol.* 1992 May 1;135(9):1019-28. | PubMed |
21. Viera AJ. Odds ratios and risk ratios: what's the difference and why does it matter? *South Med J.* 2008 Jul;101(7):730-4. | CrossRef | PubMed |
22. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol.* 1991 Jan 15;133(2):144-53. | PubMed |
23. Molina-Arias M. Estudios de casos cruzados. *Rev Pediatr Aten Primaria.* 2015 Dec;17(68):373-6. | Link |
24. Lombardi DA. The case-crossover study: a novel design in evaluating transient fatigue as a risk factor for road traffic accidents. *Sleep.* 2010 Mar;33(3):283-4. | PubMed |
25. Molina-Arias M. Diseños híbridos. *Rev Pediatr Aten Primaria.* 2016;18:8993. | Link |
26. Manterola C, Otzen T. Los Sesgos en Investigación Clínica. *Int J Morphol [Internet].* 2015 Sep;33(3):1156-64. | Link |
27. Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet.* 2002 Jan 19;359(9302):248-52. | PubMed |
28. García Marcos L, Guillén Pérez J, Orejas Rodríguez-Arango G. Epidemiología y metodología aplicada a la pediatría (V): Sesgos. *An Esp Pediatr.* 1999;50:519-524. | Link |
29. Barraza F, Arancibia M, Madrid E, Papuzinski C. General concepts in biostatistics and clinical epidemiology: Random error and systematic error. *Medwave.* 2019 Aug 27;19(7):e7687. | CrossRef | PubMed |
30. Cataldo R, Arancibia M, Stojanova J, Papuzinski C. General concepts in biostatistics and clinical epidemiology: Observational studies with cross-sectional and ecological designs. *Medwave.* 2019 Sep 25;19(8):e7698. | CrossRef | PubMed |
31. Araujo M. Confusion in clinical studies. *Medwave* 2012 May;12(4):e5349. | CrossRef |
32. Rose S, Laan MJ van der. Why match? Investigating matched case-control study designs with causal effect estimation. *Int J Biostat.* 2009 Jan 6;5(1):Article 1. | CrossRef |
33. Mansournia MA, Jewell NP, Greenland S. Case-control matching: effects, misconceptions, and recommendations. *Eur J Epidemiol.* 2018 Jan 3;33(1):5-14. | CrossRef |
34. Hidalgo B, Goodman M. Multivariate or multivariable regression? *Am J Public Health [Internet].* 2013 Jan [cited 2019 Jul 22];103(1):39-40. | CrossRef | PubMed |
35. Lu CY. Observational studies: a review of study designs, challenges and strategies to reduce confounding. *Int J Clin Pract.* 2009 May;63(5):691-7. | CrossRef |
36. Rosenbaum PR, Rubin DB. Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *J Am Stat Assoc.* 1984 Sep;79(387):516-24. | CrossRef |
37. Ghaemi S. Why you cannot believe your eyes: the three C's. In: *A clinician's guide to statistics and epidemiology in mental health: measuring truth and uncertainty.* Cambridge: Cambridge University Press; 2009.
38. Perillo MG. Choice of controls in case-control studies. *J Manipulative Physiol Ther.* 1993 Nov-Dec;16(9):578-85. | PubMed |
39. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med.* 2007 Oct 16;4(10):e296. | CrossRef | PubMed |
40. Silva Ayçaguer LC. [Causality and prediction: differences and points of contact]. *Medwave.* 2014 Sep 10;14(8):e6016. | CrossRef | PubMed |
- 41.



**Correspondence to**  
Angamos 655, Edificio R2  
Room 1107  
Reñaca, Viña del Mar  
Chile



Esta obra de Medwave está bajo una licencia Creative Commons Atribución-No Comercial 3.0 Unported. Esta licencia permite el uso, distribución y reproducción del artículo en cualquier medio, siempre y cuando se otorgue el crédito correspondiente al autor del artículo y al medio en que se publica, en este caso, Medwave.