

## REVIEW ARTICLE

### Bias in case-control studies: can it be avoided? *Viés nos estudos de caso-controle: é possível evitar?*

Fernanda de Oliveira Souza,<sup>1</sup> Samilly Silva Miranda,<sup>1</sup> Alessandra Laís Pinho Valente,<sup>1</sup> Jenny Rose Smolen,<sup>1</sup> Tânia Maria de Araújo<sup>2</sup>

<sup>1</sup>Universidade Estadual de Feira de Santana, Feira de Santana, BA, Brasil.

<sup>2</sup>University of Massachusetts, Lowell, MA, Estados Unidos.

Recebido em: 24/07/2015

Aceito em: 07/01/2016

Disponível online: 04/04/2016

nandaolisouza@gmail.com

#### DESCRITORES

Epidemiologia;  
Estudos de casos e controles;  
Viés.

#### KEYWORDS

Epidemiology;  
Case-Control Studies;  
Bias.

#### RESUMO

**Justificativa e Objetivos:** Embora os estudos de caso-controle tenham inúmeras vantagens, eles são vulneráveis a vieses que podem mascarar os verdadeiros resultados do estudo e, portanto, devem ser interpretados com cautela. O objetivo deste artigo é realizar criticamente esta abordagem, através de uma revisão da literatura, os erros que podem afetar este tipo de desenho de estudo e as possíveis estratégias para superar esses erros. **Resultados e Discussão:** Os erros mais conhecidos, presentes no delineamento de estudos caso-controle, foram os relacionados com a seleção e medição. **Conclusão:** No entanto, embora este tipo de estudo esteja sujeito a possíveis erros, as medidas preventivas postas em prática durante o planejamento de estudos caso-controle e até mesmo durante e após a sua execução podem ajudar a garantir o rigor científico. Esta revisão da literatura pode servir como um instrumento importante para o desenvolvimento e interpretação de estudos de caso-controle.

#### ABSTRACT

**Background and Objectives:** Although case-control studies have many advantages, they are vulnerable to biases that can mask the true results of the study and, therefore, should be interpreted with caution. The aim of this article is to critically perform an assessment, through a literature review, of the errors that can affect this type of study design and possible strategies to overcome these errors. **Results and Discussion:** The best known errors found in the case-control study design are those related to selection and measurements. **Conclusion:** However, although this type of study is subject to possible errors, preventive measures implemented during the planning of case-control studies and even during and after their employment can help ensure scientific rigor. This literature review can be used as an important tool for the development and interpretation of case-control studies.

## INTRODUCTION

Epidemiologic research seeks to build knowledge, yet to obtain valid and precise results in epidemiologic studies requires thoughtful choices and planning in all methodological phases. Defining the study's design is a crucial step in maintaining the necessary methodological rigor. The case-control study design has been gaining popularity in public health, but it is as with any study designs vulnerable to certain biases that must be planned or accounted for by researchers.

Janet Lane-Clayton created this study design in a pioneering 1926 study on breast cancer. Yet the design was not popular or widely accepted until after 1950, when four case-control studies about tobacco and lung cancer were published. Those landmark studies, which identified tobacco as a risk factor in lung cancer, were convincing and very influential. While the case-control study design has since been modified and improved upon, those studies laid the foundation for the rise in popularity of the case-control study.<sup>1</sup>

Case-control studies are observational epidemiological studies that are longitudinal, retrospective, and analytic. They were created to investigate etiologic associations in diseases of low incidence, or those with a prolonged latency period. The defining feature of this design is that it compares a group of people who have a defined disease, and a group of those who do not, and assesses the frequency of past exposure to the risk factors of interest.<sup>2,3</sup>

In this type of study, those people who have the disease or attribute of interest to the study are considered the case group. This case group is compared to a group of people who do not have the disease or attribute of interest, considered the control group.<sup>4</sup> Song et al. (2010) point out that in comparison to cohort studies, case-control studies are quick, relatively inexpensive, and demand comparatively fewer participants. They also allow investigators to evaluate multiple exposures and diverse risk factors within the same study.<sup>5</sup>

However, in choosing a study design, the risks as well as the benefits should be considered. Although case-control studies present numerous advantages, they also have an increased vulnerability to certain types of bias.<sup>3</sup> This article reviews the literature to critically assess the biases that affect this type of study design, as well as present possible methods of overcoming or minimizing these biases to assist researchers who choose this type of epidemiologic study design.

## METHODS

To achieve this objective, a bibliographic search was performed in PubMed, Scielo, and Google Scholar about bias in case control studies, as well as bias in epidemiologic studies. We searched in relevant Epidemiology textbooks as well. We did not restrict the search by year. This search was conducted from April to July 2015.

## RESULTS E DISCUSSION

### ERRORS IN EPIDEMIOLOGIC STUDIES

The field of epidemiology has sought to understand the factors that may interfere in testing hypotheses, among them errors that can mask the true results of a study. Before presenting the errors commonly found in case-control studies, a basic understanding of the concepts and types of known error such as random and systematic errors is necessary.

Random error occurs when the value measured in the sample diverges, randomly, from the true value of the population. Random error can occur in any study design, and is not particular to case-control studies. Using a larger sample size reduces random error, and this type of error can be estimated by statistical tests. The second type of error, known as systematic error or bias, is defined as any process, in any stage of research, in which incorrect methodology over the course of the investigation distorts the result.<sup>6-8</sup>

Since bias includes any and all distortion during the process of investigation, it can occur in any kind of study design. However, there are certain kinds of bias to which case-control studies are vulnerable. The types of bias can be classified in the following categories: selection bias, information bias, and confounding.<sup>9</sup> These categories will be explored in the present article in the context of the case-control study, along with possible strategies researchers can use to overcome them.

### THE BIASES OF CASE-CONTROL STUDIES

Case-control studies begin with the effect (disease) in order to investigate the cause (exposure). This structure confers both the strengths and the weaknesses of this design. On the other hand, case-control studies may be subject to certain biases.<sup>10</sup>

### SELECTION BIAS

Selection bias occurs when the procedures that were used to select the participants, and/or factors that influenced participation in the study distort the results.<sup>8</sup> Bayona & Olsen (2004) state that this occurs because the structure of the sample is sufficiently different from the target population the sampling strategy did not result in a sample that mirrors the structure of the target population.<sup>11</sup>

In case-control studies, this selection bias can occur when the controls are not representative of the base population that produced the cases. The controls should not be selected depending on their level of exposure, but rather should represent the distribution of the exposure in the base population.<sup>8</sup>

Within the umbrella of selection bias, there are other specific kinds of bias to which case-control studies are vulnerable. Participation or response bias (also called non-response or self-selection bias) occurs when those who consent to participate in the study (and thus

respond to the survey) are different from those who refuse participation, compromising the representativity of the study. In general, those that participate in research studies have greater availability, better quality of life, are more concerned with their health, and take better care of their health. For a case-control study examining an exposure such as soda consumption, this could mean that those who choose to participate as controls are healthier than the average person, and therefore drink less soda; in this case selection bias could lead to overestimation of the association between the exposure and disease.<sup>8,11</sup>

It is important to consider strategies to minimize selection bias in the planning stage of a study, before any recruitment or execution of data collection. One method of verifying the proper selection of controls is checking that if a control were to present the disease during the study, they could be a case. If a control could end up as a case, this implies an adequate control group.<sup>9,12</sup> Another effective strategy, if used properly, is to match cases with controls. This is almost always accomplished by pairing of demographic or social characteristics, or even by place of residence. Matching should take into consideration the possible confounding variables this strategy controls for the association between the confounding variable(s) and the disease, making it possible to minimize the influence of confounders, and also improves the efficiency of the study.<sup>5</sup>

Although less frequent, selection bias can also occur in the selection of cases. This happens when the exposed cases have a greater possibility of being selected than the non-exposed cases (or vice versa). According to Rêgo (2010), the inclusion criteria to determine cases must be defined in a way that assures that all the true cases have an equal probability of entering into the group. This should also ensure that no false case is selected, therefore avoiding the possibility that false cases distort the estimate of the measure of association in the direction of the null hypothesis.<sup>2</sup>

Giannini et al. (2012), in their case-control study on voice disorders and stress in teachers, reduced possible selection bias by opting to draw from professors who lectured in the same schools to create the control group. This guaranteed the maximum similarity with the case group, with the same probability of exposure to physical, chemical, and biological risks in the school environment.<sup>13</sup> These authors considered the limitations of case-control studies not only when trying to control bias, but also when clarifying to the reader that the main methodological difficulty they encountered was how to conceptualize and define what was considered a case. Although voice disorders are complicated to diagnose, the researchers recognized that a clear case definition is crucially important in case-control studies. Ideally, case controls studies should use highly sensitive and specific laboratory exams, complemented by clinical diagnosis. If that is not possible, surveys that have been previously validated and can clearly discern cases from non-cases may be used.

Berkson Bias is another form of selection bias, one which applies to the use of hospital controls. It points out that the relative frequency of the diseases and etiologic

exposure factors observed in a hospitalized population may be biased when compared to a community population for which the results are referenced. This bias occurs because of the different probabilities of hospitalization, with differences both in relation to the disease itself and in relation to the factors of exposure, and even the possibility of the combination of diseases and exposure factors.<sup>14</sup>

If the population base that originated the case is unknown, the nature of the disease in question must be considered when defining a control for a hospital case. One possible method could be to define controls as those individuals who also entered the hospital with a disease of a similar severity to the case. Another way to minimize bias in the selection of controls would be to use more than one control group, although this does risk finding discrepant results.<sup>9</sup>

Another type of bias that can occur in this study design is the selection of prevalent cases instead of incident cases, known as Neyman's bias or the incidence-prevalence bias. Prevalence is affected by the duration of the disease, which in turn is influenced by treatments and cures, as well as the mortality associated with the disease. For example, a case-control study could be carried out to study the relationship between cigar smoking and stroke, where all cases are interviewed within a month of having the stroke. If those who smoke cigars die more frequently, then the surviving cases will have a lower frequency of cigar smoking, thus decreasing the association between cigars and stroke.<sup>15</sup> This bias can be avoided by favoring the selection of incident cases, and if this is not possible, authors should discuss Neyman's bias as a limitation that could affect the association studied.

## INFORMATION BIAS

Information bias refers to distortions in the effect estimates obtained in epidemiologic studies that are due to errors in measurement of the exposure or outcome of interest. Data can be collected in a variety of ways: in-person, by telephone, through mailed interviews, etc. No matter the method of data collection, it is absolutely essential to obtain accurate information as much for the cases as for the controls. By nature, case-control studies require that some individuals are selected because they present with the disease, and therefore case control studies are particularly influenced by information bias the fact that an individual has or does not have the disease being studied can affect accurately they recall information about the exposure (or lack of exposure) to the factor of interest.<sup>9</sup>

All retrospective studies are susceptible to this memory bias, since by nature they depend on memory. Generally, cases tend to have a better recall than controls about the object of study, which could influence responses given to certain questions. Cases tend to better search their memories to identify what could have caused their disease; healthy controls do not have the same motivation.<sup>16</sup>

Considering that the capacity to remember information may be different between the cases and the controls, researchers should strategize how to avoid this bias. Oliveira and Parente (2010) suggest that the best form of

minimizing memory bias in case-control studies is to use data that was collected in a systematic way before the development of the disease; for example, organized and complete medical records could be used.<sup>7</sup>

### OBSERVER BIAS

Another bias to be considered is that of the observer. This occurs when the way the observer obtains information differs if the individual is a case or a control, so that knowledge of the stage of disease or stage of exposure can influence the level of detail in the data collection. Even if the interviewers are administering a standardized survey, if they emphasize questions differently or administer the survey differently for different groups of interviewees (cases vs. controls; by stage of disease or exposure) this systematic difference in data collection could bias the results.<sup>15</sup>

To control observer bias, it is important that observations be made under the same conditions in both the case and control groups. The strongest strategy to avoid observer bias is for the investigator to be blinded as to whether the subject has the disease or not (and therefore whether the subject is part of the case or control group) in order to avoid possible bias in the data collection. If blinding is not possible, this bias can be minimized by rigorously training the interviewers in a standardized manner so that all interviewers follow the same exact data collection process for cases and controls.

### CONFOUNDING

Confounding is also a systematic error, but not due to any methodological error during the study. It is present when there is a variable that is associated with the exposure and associated with the outcome, but is not directly in the causal pathway between the two. For example, in a case-control study of smoking and lung cancer, men and women may have different rates of smoking. Therefore, gender could be considered a confounding factor, and if there are different proportions of men in the case group as compared to the control group it could affect the association the study finds between smoking and lung cancer. There are several different methods of minimizing confounding.

According to Grimes and Schulz (2002), control for confounding can occur before or after fieldwork.<sup>16</sup> Strategies to reduce confounding include restricting the sample through inclusion/exclusion criteria, stratification, or pairing. However, the advantages and disadvantages of these strategies should be weighed. For example, while restricting the sample can avoid confounding, it can also limit external validity if the study on smoking and lung cancer was restricted to men, the results will only be generalizable to men.

Case-control studies are subject to numerous biases, but these biases can be controlled or minimized. It is essential for the researcher to be aware of the biases that can occur in a study, and above all, describe clearly the limitations and challenges as well as the mechanisms used to control them (Chart 1).

Goi et al. (2014), in analyzing whether case-control studies in psychiatry published in Brazilian journals adequately described how they controlled for biases, observed that only 12.5% of the 16 studies assessed fulfilled this requirement. This shows how necessary it is for the researcher to choose a methodological design befitting the proposed objective while considering the existence of possible biases and the methods of controlling them, and present this information for the reader.<sup>17,18</sup>

**Chart 1.** Summary of the principal biases encountered in case-control studies, and the principal strategies to overcome these biases.

Principal biases of case-control studies	Strategies to account for these biases
<b>Selection</b>	
Bias in selection of controls	Verify that the controls could also be selected as cases if they were to present with disease in study; Pair cases with controls / use confounding variables to pair
Bias in selection of cases	Assure that all true cases have an equal probability of entering the group
Berkson's bias  Neyman's bias	Define controls by assessing if these individuals were admitted to the hospital with a disease of similar severity as the case  Count only incident cases, and if this is not possible describe the limitation of the use of prevalent cases for the outcome of interest
<b>Measurement</b>	
Recall bias	Use data that was collected in a systematic manner before the onset of disease
Observer bias	Select controls and assure that all observations will be performed in both groups under the same conditions. Train and standardize the procedures and behaviors of the observer.
<b>Confounding</b>	Assure comparability between the groups of the study; Pair; Stratify

### CONSIDERATIONS

Case-control studies are investigations that, as with any study design, present both advantages and disadvantages. The present article points out the principal disadvantages known in the literature, and the principal biases that commonly occur in this kind of study design. It should be observed that the biases most widely recognized in case-control studies are those related to selection and measurement bias. Although this study design has the possibility of errors, there are strategies which can be used in the planning of the study, as well as during and even after its execution to guarantee scientific rigor, above all because case-control have a recognized scientific importance.

## CONFLICT OF INTEREST AND SOURCE OF FUNDING STATEMENT:

The authors declare that they have no conflicts of interests in this Guest Editorial.

## REFERENCES

1. 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;47:359–365.
2. Rêgo MAV. Estudos caso estudos caso-controle: uma breve revisão. *Gaz Méd Bahia* 2010;80(1):101-110.
3. Rouquayrol MZ. *Epidemiologia e Saúde*. 6a ed. Rio de Janeiro: MEDSI; 2003, p. 728.
4. Sesso RCC, Filho AC, Marcopito LF, et al. Avaliação do estudo tipo caso-controle na pesquisa médica. *Rev Paul Med* 1987;105(2):96-99.
5. Song JW, Chung KC. Observational Studies: Cohort and Case-Control Studies. *Plast Reconstr Surg* 2010;126(6):2234–2242. doi: 10.1097/PRS.0b013e3181f44abc.
6. Coutinho M. Princípios de Epidemiologia Clínica Aplicada à Cardiologia. *Arq Bras Cardiol* 1998;71(2):109-116. doi: 10.1590/S0066-782X1998000800003.
7. Bonita R, Beaglehole R, Kjellström T. *Epidemiologia básica* [Internet]. 2.ed. Organização Mundial da Saúde; São Paulo: Santos; 2010. 213 p. [citado em 2015 mai 30]. Disponível em: [http://whqlibdoc.who.int/publications/2010/9788572888394\\_por.pdf](http://whqlibdoc.who.int/publications/2010/9788572888394_por.pdf).
8. Botelho F, Silva C, Cruz F. Epidemiologia explicada – Viéses. *Acta Urologica* 2010;3:47–52.
9. Rodrigues LC, Werneck GL. Estudos Caso-controle. In: Medronho R, Bloch KV, Luiz RR, Werneck GL. São Paulo: Epidemiologia. Atheneu; 2009.
10. Lima-Costa MF, Barreto SM. Tipos de estudos epidemiológicos: conceitos básicos e aplicações na área do envelhecimento. *Epidemiol Serv Saúde* 2003;12(4):1679-4974. doi: 10.5123/S1679-49742003000400003.
11. Bayona M, Olsen C. Observational Studies and Bias in Epidemiology [Internet]. College Entrance Examination Board. 2004. [citado em 2015 jun 02]. Disponível em: [http://www.collegeboard.com/prod\\_downloads/yes/4297\\_MODULE\\_19.pdf](http://www.collegeboard.com/prod_downloads/yes/4297_MODULE_19.pdf).
12. Kopec JA, Esdaile JM. Bias in case-control studies. A review *J Epidemiol Health* 1990;44(3):179-186. doi: 10.1002/0470011815.b2a03013.
13. Giannini SPP, Latorre MRDO, Ferreira LP. Distúrbio de voz e estresse no trabalho docente: um estudo caso-controle. *Cad Saúde Pública* 2012;28(11):2115-2124. doi: 10.1590/S0102-311X2012001100011.
14. Feinstein AR, Walter SD, Horwitz RI. An analysis of Berkson's bias in case-control studies. *J Chronl* 1986;39(7):495-504.
15. Delgado-Rodriguez M, Llorca J. Bias. *J Epidemiol Community Health* 2004;58:635–641. doi: 10.1136/jech.2003.008466
16. Grimes DA, Schulz KF. Bias and causal associations in observational research. *The lancet* 2002;359(9302):248-252. doi: 10.1016/S0140-6736(02)07451-2.
17. Oliveira MAP, Parente RCM. Estudos de Coorte e de Caso-Controle na Era da Medicina Baseada em Evidência. *Bras J Video-Sur* 2010;3(3):115-125.
18. Goii PD, GoiiJD, Cordinii KL, et al. Evaluating psychiatric case-control studies using the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement. *Med J* 2014;132(3):178-83. doi: 10.1590/1516-3180.2014.1323653.