

General concepts in biostatistics and clinical epidemiology: Observational studies with cross-sectional and ecological designs

Ricardo Cataldo^a, Marcelo Arancibia^{a,b}, Jana Stojanova^{a,b}, Cristian Papuzinski^{a,b,*}

^a Cátedra de Metodología de la Investigación Científica, Escuela de Medicina, Universidad de Valparaíso, Viña del Mar, Chile

^b Centro Interdisciplinario de Estudios en Salud (CIESAL), Universidad de Valparaíso, Valparaíso, Chile

*Corresponding author cristian.papuzinski@uv.cl

Citation 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;19(8):e7698

Doi 10.5867/medwave.2019.08.7698

Submission date 27/6/2019

Acceptance date 12/9/2019

Publication date 25/9/2019

Origen 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

Tipo de revisión Externally peer-reviewed by three reviewers, double-blind.

Key words observational study, cross-sectional studies, epidemiology, biostatistics, bias

Abstract

Observational studies evaluate variables of interest in a sample or a population, without intervening in them. They can be descriptive if they focus on the description of variables, or analytical when comparison between groups is made to establish associations through statistical inference. Cross-sectional studies and ecological—also called correlational—studies are two observational methodological designs. Cross-sectional studies collect the data of the exposure variable and the outcome at the same time, to describe characteristics of the sample or to study associations. Ecological studies describe and analyze correlations among different variables, and the unit of analysis is aggregated data from multiple individuals. In both types of studies, associations of interest for biomedical research can be established, but no causal relationships should be inferred. This is the second 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. In this review, we address general theoretical concepts about cross-sectional and ecological studies, including applications, measures of association, advantages, disadvantages, and reporting guidelines. Finally, we discuss some concepts about observational designs relevant to undergraduate and graduate students of health sciences.

Key ideas

- Cross-sectional designs collect study variables simultaneously, and the unit of analysis is the individual. They are useful in determining the prevalence and facilitate rapidly establishing associations among variables
- Ecological studies analyze correlations among variables whose unit of analysis is grouped data. They are usually easy to conduct and allow the study of large populations.
- These observational studies cannot establish causal inferences but do permit establishing statistical relationships of great importance for biomedical research and public health.

Introduction

An essential classification in clinical epidemiology is based on the criterion of observation versus experimentation, that is, if researchers focus on the observation of measured variables or if they apply an

intervention among study participants. In the first case, we refer to observational studies, where data of interest are collected and then analyzed descriptively and/or analytically, which includes the application of interviews, measuring instruments, laboratory tests, among others, but without intervening the exposure variable. In the second

case, researchers handle the exposure variable, which involves subjecting participants to a controlled intervention to study the modification of some estimators of interest (the outcome or response variable). It is in a sense a clinical experiment, which in clinical epidemiology is called a clinical trial. Today, observational studies play an essential role in various aspects of health science research and even provide answers when clinical trials are ethically questionable or difficult to perform.

This review is the second release of a methodological series of six narrative reviews about general topics in biostatistics and clinical epidemiology. Each article will cover one of six topics based on content from publications available in the main databases of scientific literature and 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. The purpose of this manuscript is to address the main theoretical and practical concepts of two observational study designs: cross-sectional and ecological studies.

Descriptive studies versus analytical studies

Another classification in the taxonomy of methodological designs is the definition of studies as descriptive and/or analytical. Studies have a descriptive purpose if their objective is merely to describe the frequency distribution of the variables without the pretense of obtaining conclusions about associations¹, or analytical if they incorporate some level of inferential statistical analysis with the purpose of establishing associations from the data. Descriptive studies constitute a large part of published research and have contributed to the understanding of the semiology and natural history of diseases, the frequency of certain phenomena in the population, the study of infrequent conditions and the establishment of interventions, giving rise to the origin of new hypotheses. Among the descriptive studies, we find case reports and case series, where infrequent conditions are presented at the level of diagnosis, treatment and/or prognosis². These used to be the first source of evidence regarding emerging conditions, such as the clinical observation of blindness in newborns that led to the association with high concentrations of oxygen in incubators, or hepatocellular adenoma in young women, concluding the relationship with exposure to high doses of contraceptive drugs¹. In case reports or case series, a descriptive analysis of the reported data is presented³. Various authors place cross-sectional studies (studies in individuals) and ecological studies (studies in population) within the category of descriptive studies. However, both designs can have an analytical orientation, where hypothesis tests are applied using at least two groups of participants (comparison groups) to obtain statistical inference; therefore, they can also be classified as analytical studies³⁻⁵.

Cross-sectional studies

The central element of cross-sectional studies is that both the variable considered an exposure (variable X, independent, explanatory, predictive or factor) and the outcome variable (variable Y, dependent, explained, predicted or response) are measured simultaneously, that is, temporality is cross-sectional or in a single moment. This

does not permit ensuring that the exposure has preceded the outcome because there is no follow-up over time. In cross-sectional studies, a representative sample of a larger population can be studied, or an entire population can be analyzed, such as with a census. In both situations, frequency or prevalence of a condition of interest can be determined, a reason why these are also known as “prevalence studies.” This could include a pathology, a characteristic, a factor conceptualized in the literature as prognostic, such as a protective factor or a risk factor, among others. However, the association between two variables of interest can also be studied, thus exhibiting an analytical orientation^{3,5}. A cross-sectional study is exemplified in the following example⁶.

Example 1. A study sought to determine the prevalence of asthma in children and analyze its association with being a passive smoker, being exposed to vehicular traffic (both risk factors) and the intake of dehydrated fruit (a possible protective factor). The researchers found that the prevalence of asthma increased with the number of smokers with whom they lived, but it was not associated with living near the main avenue or the consumption of dehydrated fruits. Thus, in this cross-sectional study, there is both a descriptive (an estimate of prevalence) and an analytical component (study of the associations between the variables).

Measures of association

Although in the previous example it was possible to establish the associations using advanced statistical methods, it would not be possible to directly determine the risk as this is reserved for studies that have a longitudinal temporal approach⁷; it is thus a matter of methodological design and not statistical analysis. Therefore, the appropriate association measures in the case of cross-sectional studies are the odds ratio (OR) and the prevalence ratio (PR). The odds ratio can be defined as the excess or reduction in the advantage that exposed individuals have in presenting the condition compared to not presenting it, concerning the advantage (or reduction) in non-exposed individuals presenting the condition compared to not presenting it. For its part, the interpretation of the prevalence ratio is simpler, more direct and to some degree intuitive, since it indicates how many times individuals exposed to a phenomenon are more likely to present the condition with respect to those not exposed⁸⁻¹⁰. Although they correspond to different concepts, interpreting the odds ratio as a prevalence ratio is a conceptual error frequently observed in published research.

A particular type of cross-sectional study is a diagnostic test study, where the ability of a test to discriminate between the presence and absence of disease (index test) is evaluated for the purpose of diagnosing a disease¹¹. It is usually performed by comparing the test results with a reference standard (also known as the gold standard or truth criterion) in healthy and those with the condition, to later apply in people suspected to have the disease¹². These studies evaluate the operational characteristics of the index test, such as its specificity, sensitivity, predictive values and likelihood ratios¹³. Example 2 presents a diagnostic test study, whose design corresponds to a cross-sectional study¹⁴.

Example 2. A cross-sectional study analyzed the diagnostic utility of a rapid antigen test (index test) for the diagnosis of acute tonsillitis in children between 2 and 14 years. This test was compared with pharyngeal culture, considered as the standard diagnostic reference. A sensitivity of 86.5% and a specificity of 91.5% were found, demonstrating that the test is useful for the diagnosis of the pathology in this context.

Advantages and disadvantages

Cross-sectional studies are usually quick to execute. Because they do not involve temporal follow-up, loss of follow-up is not a problem, and associated economic costs are lower, allowing associations to be established quickly¹. The main disadvantage is the issue of temporality since it is not clear that the exposure variable (cause) precedes the result variable (effect) and it is not possible to establish a causal relationship^{1,15}; thus results must be interpreted prudently and in context. Likewise, this design is not very useful in infrequent pathologies or those where prevalence changes rapidly, as in the case of infectious diseases⁵.

Ecological or correlational studies

Ecological or correlational studies share the central characteristic of cross-sectional studies, since, regarding temporality, both explanatory and explained variables are collected simultaneously. They are known as "ecological" as investigations of this type use geographical areas to define the units of analysis. Indeed, their particularity lies in the unit of analysis: grouped data are analyzed (ecological units), corresponding to estimators determined from summaries of individual data; thus they are studies based on populations¹⁶. The frequency of a condition in a population is studied, and its correlation (hence the name "correlational" studies) with one or more exposure variables that are also measured in aggregate⁵. For example, an ecological study¹⁷ analyzed the inequality in the distribution of otolaryngologists in Latin American countries, concluding that in all countries specialists were more frequently found in socio-geographically advantageous areas and capital cities, demonstrating high inequality in distribution; the authors emphasize the importance of implementing policies that improve access to this medical discipline.

Some of its advantages include the mapping of diseases and their risk factors, the realization of large-scale comparisons, and the study of public health strategies^{16,18}. Likewise, ecological studies have contributed significantly to the analysis of occupational exposures to harmful agents, as in the case of the association between exposure to asbestos and occurrence of mesothelioma^{18,19}.

Although the main type of ecological study is the geographical one, where a condition of interest is compared between geographic regions, it is also possible to monitor a population over time to evaluate its changes, as in the case of longitudinal ecological studies. These are particularly sensitive to biases, such as those associated with the method of disease determination, as examinations and diagnostic criteria tend to improve over time. Other types of ecological study are studies of migrant populations, which are used to discriminate genetic factors from environmental factors based on geographical and cultural variation. Nonetheless, it should be taken into account

that the migrant population may not be representative of the population of origin and that health may be affected by the migration process itself. Example 3 shows an ecological study in migrant populations^{20,21}.

Example 3. In a study by Ødegaard published in 1932, titled "Emigration and insanity," it was observed that the rate of hospitalization (ecological unit) for schizophrenia was higher in cases that had emigrated to the United States compared to its compatriots residing in Norway, which opened the debate about the role that environmental factors play in the psychopathology of psychosis. However, the results should be interpreted with caution for the reasons discussed.

Measures of association

The measure of association in these studies is a correlation coefficient (hence the name "correlational studies") that indicates the degree of a linear association between two variables that are conceptualized as exposure and outcome¹. The study of variables associated with the dependent variable, analysis of confounding variables and the construction of predictive models for the response variable could be considered using multivariate statistical regression methods²².

Advantages and disadvantages

In general, ecological studies are easy to conduct, since data is usually already collected in statistics from public institutions, or open-access registries such as national surveys²³. This would also solve the bioethical complexity linked to direct study in humans and its economic cost¹. Also, they facilitate the study of large populations.

The primary disadvantage associated with inference from ecological studies is related to the reduction of information that may occur in the process of aggregating data, which does not permit identifying associations at an individual level¹⁶. As data is analyzed in aggregate form, the relationship between exposure and outcome cannot be empirically determined at the individual level, so to infer about causal mechanisms at an individual level from aggregate statistics of the group in which an individual belongs (for example, the hospitalization rate of a country) is an error known as ecological fallacy, ecological bias or fallacy of division^{1,18}. For example, one study²⁴ demonstrated a very significant linear correlation between the consumption of chocolate per capita and the number of Nobel prizes for every 10 million people in 23 countries studied ($r = 0.791$, $p < 0.0001$); however, this does not ensure that award-winners consumed large amounts of chocolate. Another disadvantage, typical of studies in which the variables of interest are measured at the same time, is temporal ambiguity since it is not possible to define which phenomenon occurred first. Finally, statistical analysis of these designs could be hindered by multicollinearity, a phenomenon where there is a correlation between predictive (independent) variables of a multivariate model, which could reduce the relevance of variables of greater interest²⁵.

Reporting guidelines

In 2007, an international collaboration of epidemiologists, methodologists, statisticians, researchers, and journal editors published the

Strengthening the Reporting of Observational Studies in Epidemiology, or STROBE reporting guideline (<http://www.strobe-statement.org>)²⁶, based on the experience with the CONSORT guideline which guides reporting for randomized controlled trials²⁷. Its purpose is to promote the clear and transparent reporting of research and is therefore not a quality assessment tool. STROBE focuses on the three most widespread observational methodological designs: cross-sectional studies, case-control studies, and cohort studies. It includes twenty-two items grouped into six domains: title and summary, introduction, methods, results, discussion and additional information^{27,28}. Although the use of reporting guidelines has been emphasized internationally, the use of STROBE is not homogeneous in the published literature^{29,30}. There is currently no similar initiative for ecological studies.

Preventing and controlling confounding

A fundamental challenge for observational studies is the prevention and control of potential biases that may threaten their internal validity, especially confounding. Confounding can occur, for example, when the groups compared differ in baseline characteristics (such as biodemographic characteristics), such that there are intergroup differences in addition to the variable of interest³¹. Many observational studies use data that were originally collected for purposes other than research objectives, for example, national surveys, hospital statistics, among others; this represents another source of confounding. To respond to these concerns, at the level of design for a cross-sectional study, strategies such as the application of rigorous eligibility criteria or the restriction can be used (for example, strict selection of subjects who present the characteristic to be “neutralized,” or selecting those in whom it is absent)³². At the level of statistical analysis, a stratified analysis can be employed, which is the analysis according to strata of individuals grouped according to a confounding variable, such as age and sex. As mentioned, multivariate statistical regression models can be used, whose purpose is the identification of the variables that, when adjusting the model, act as confounding variables³³. Ways of controlling confounding at the level of data analysis will be elaborated further in the next article in this series.

Final considerations

Although they are usually known as prevalence studies that primarily suggest a descriptive purpose, cross-sectional studies often lead to the study of associations when a comparison group is available. If the primary objective is to determine the prevalence of a condition, the appropriate design is a cross-sectional study. However, sampling must be random; non-probabilistic sampling only permits the study of frequency. In the study cited in Example 1, random sampling was carried out in different schools in the United Kingdom to determine the prevalence of asthma in children⁶. The study of prevalence should not be confused with that of incidence. The determination

of the incidence (the frequency of outcomes in a given period) is performed in cohort studies (observational designs whose temporal axis is longitudinal, regardless of whether data is collected prospectively or retrospectively).

Some authors have pointed out that due to phenomena that have a great influence on the results, such as the ecological fallacy, ecological studies should only be undertaken when it is not possible to perform an analysis of the individual data³¹. However, due to the advantages and opportunities mentioned, they are often the first step, especially for public health objectives, such as an analysis of the geographic distribution of specialists in otolaryngology¹⁷ or environmental factors in psychosis²⁰.

Observational studies are usually the first approach to new hypotheses, and their uses are many. They may help to identify statistical hypotheses that can later be studied through hypothesis testing, giving rise to associations. Cross-sectional and ecological studies, due to their temporality, do not allow causal hypotheses to be established. They must be conducted rigorously, considering that they are vulnerable to multiple biases, especially confounding, which can be prevented at the level of design, and controlled during the statistical analysis. As a whole, observational studies offer the possibility for new ways of looking at things (Figure 1).

Notes

Roles and contributions of authorship

MA, JS, and CP 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.

RC, MA, CP: conceptualization, methodology, investigation, resources, writing (original draft preparation), writing (review and editing), visualization, supervision, project administration.

JS: conceptualization, methodology, investigation, resources, writing (original draft preparation), writing (review and editing), visualization.

Funding

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

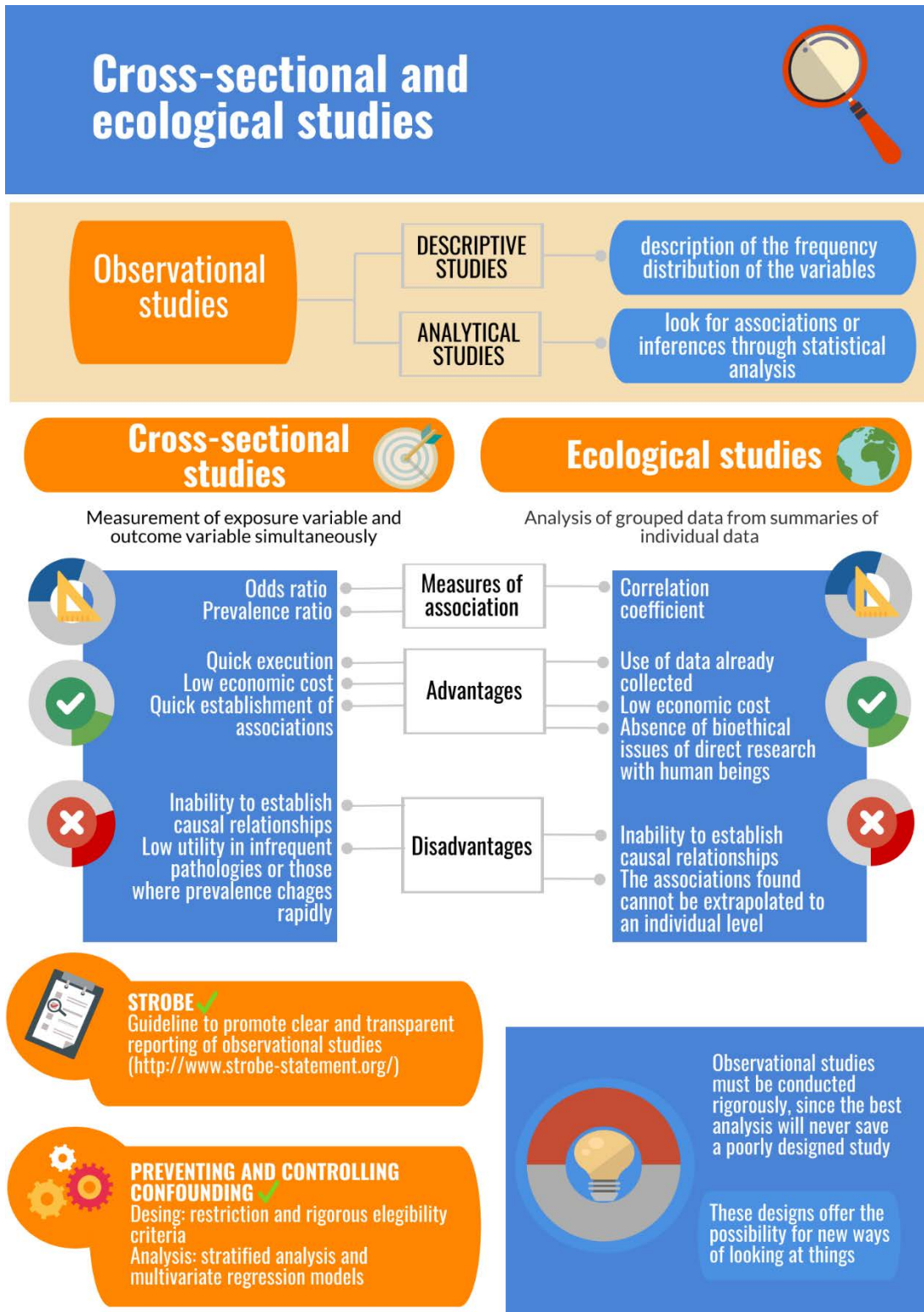
Competing interests

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.

Language of submission

Spanish. The English translation of the originally submitted article has been copyediting by the Journal.

Figure 1. Infograph of cross-sectional and ecological studies. Source: designed by the authors.



References

1. Grimes DA, Schulz KF. Descriptive studies: what they can and cannot do. *Lancet*. 2002 Jan 12;359(9301):145-9. | PubMed |
2. Sayre JW, Toklu HZ, Ye F, Mazza J, Yale S. Case Reports, Case Series - From Clinical Practice to Evidence-Based Medicine in Graduate Medical Education. *Cureus*. 2017 Aug 7;9(8):e1546. | CrossRef | PubMed |
3. Araujo M. General categories of clinical studies. *Medwave* 2011 Feb;11(2):e4875. | CrossRef |
4. Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. *Lancet* (London, England). 2002 Jan 5 [cited 2019 Apr 30];359(9300):57-61. | Link |
5. Aggarwal R, Ranganathan P. Study designs: Part 2 - Descriptive studies. *Perspect Clin Res*. 2019 Jan-Mar;10(1):34-36. | CrossRef | PubMed |
6. Lewis SA, Antoniak M, Venn AJ, Davies L, Goodwin A, Salfield N, et al. Secondhand smoke, dietary fruit intake, road traffic exposures, and the prevalence of asthma: a cross-sectional study in young children. *Am J Epidemiol*. 2005 Mar 1;161(5):406-11. | Link |
7. Araujo M. The temporality of clinical trials. *Medwave* 2011 May;11(05):e5020. | CrossRef |
8. Martínez-González MA, de Irala-Estevez J, Guillén-Grima F. [What is an odds ratio?]. *Med Clin (Barc)*. 1999 Mar 27;112(11):416-22. | PubMed |
9. Lee J, Chia KS. Use of the prevalence ratio v the prevalence odds ratio as a measure of risk in cross sectional studies. *Occup Environ Med*. 1994 Dec;51(12):841. | PubMed |
10. Schiaffino A, Rodríguez M, Psarín M, Regidor E, Borrell C, Fernández E. ¿Odds ratio o razón de proporciones? Su utilización en estudios transversales. *Gac Sanit*. 2003;17(1):70-4. | Link |
11. Knottnerus JA, Muris JW. Assessment of the accuracy of diagnostic tests: the cross-sectional study. *J Clin Epidemiol*. 2003 Nov;56(11):1118-28. | PubMed |
12. Araujo M. Diagnostic clinical trials. *Medwave* 2011 Jul;11(07):e5067. | CrossRef |
13. Kumar R. Evaluation of diagnostic tests. *Clin Epidemiol Glob Heal*. 2016 Jun 1;4(2):76-9. | Link |
14. Regueras De Lorenzo G, Santos Rodríguez PM, Villa Bajo L, Pérez Guirado A, Arbesú Fernández E, Barreiro Hurlé L, et al. [Use of the rapid antigen technique in the diagnosis of *Streptococcus pyogenes* pharyngotonsillitis]. *An Pediatr (Barc)*. 2012 Sep;77(3):193-9. | CrossRef | PubMed |
15. Sedgwick P. Cross sectional studies: advantages and disadvantages. *BMJ*. 2014 Mar 26;348(mar26 2):g2276-g2276. | Link |
16. Wakefield J. Ecologic studies revisited. *Annu Rev Public Health*. 2008;29:75-90. | PubMed |
17. Bright T, Mújica OJ, Ramke J, Moreno CM, Der C, Melendez A, et al. Inequality in the distribution of ear, nose and throat specialists in 15 Latin American countries: an ecological study. *BMJ Open*. 2019 Jul 19;9(7):e030220. | CrossRef | PubMed |
18. Sedgwick P. Ecological studies: advantages and disadvantages. *BMJ*. 2014 May 2;348:g2979. | CrossRef | PubMed |
19. Coggon D, Geoffrey R, Barker D. Ecological studies. In: *Epidemiology for the uninitiated*. 5th ed. London: BMJ Books; 2003.
20. Ødegaard Ø. Emigration and insanity. *Acta Psychiatr Neurol Scand Suppl*. 1932;4:206.
21. Tarricone I, Tosato S, Cianconi P, Braca M, Fiorillo A, Valmaggia L, et al. Migration history, minorities status and risk of psychosis: an epidemiological explanation and a psychopathological insight Background: psychosis, migration and minorities. Vol. 21, *Society and psychopathology Journal of Psychopathology*. 2015. | Link |
22. Alexopoulos EC. Introduction to multivariate regression analysis. *Hippokratia*. 2010 Dec;14(Suppl 1):23-8. | PubMed |
23. Saunders C, Abel G. Ecological studies: use with caution. *Br J Gen Pract*. 2014 Feb;64(619):65-6. | CrossRef | PubMed |
24. Messerli FH. Chocolate consumption, cognitive function, and Nobel laureates. *N Engl J Med*. 2012;18:1562-4.
25. Vatcheva KP, Lee M, McCormick JB, Rahbar MH. Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. *Epidemiology (Sunnyvale)*. 2016 Apr;6(2). pii: 227. | PubMed |
26. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med*. 2007 Oct 16;4(10):e296. | PubMed |
27. Cartes-Velasquez R, Moraga J. Pautas de chequeo, parte III: STROBE y ARRIVE. *Rev Chil Cirugía*. 2016 Sep;68(5):394-9. | Link |
28. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Int J Surg*. 2014 Dec;12(12):1500-24. | CrossRef | PubMed |
29. Mannocci A, Saule R, Colamesta V, D'Aguanno S, Giraldi G, Maffongelli E, et al. What is the impact of reporting guidelines on Public Health journals in Europe? The case of STROBE, CONSORT and PRISMA. *J Public Health (Oxf)* [Internet]. 2015 Dec 23;37(4):737-40. | Link |
30. Pouwels KB, Widyakusuma NN, Groenwold RHH, Hak E. Quality of reporting of confounding remained suboptimal after the STROBE guideline. *J Clin Epidemiol*. 2016 Jan;69:217-24. | Link |
31. 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. | Link |
32. Araujo M. Confusion in clinical studies. *Medwave* 2012 May;12(4):e5349. | CrossRef |
33. Hidalgo B, Goodman M. Multivariate or multivariable regression? *Am J Public Health*. 2013 Jan;103(1):39-40. | CrossRef | PubMed |

Correspondence to
Angamos 655, Edificio R2
Oficina 1124
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.