RBSO Revista Brasileira de Saúde Ocupacional
ISSN: 2317-6369 (online)

Ronir Raggio Luiz[a]
https://orcid.org/0000-0002-7784-9905

[a] Universidade Federal do Rio de Janeiro, Instituto de Estudos em Saúde Coletiva. Rio de Janeiro, RJ, Brazil.

Contact:
Ronir Raggio Luiz
E-mail:
ronir@iesc.ufrj.br

How to cite (Vancouver):
Luiz, RR. Statistical inference versus causal inference. Rev Bras Saude Ocup [Internet]. 2025;50:edisfl1. Available from: https://doi.org/10.1590/2317-6369/00124en2025v50edisfl1

Statistical inference *versus* causal inference

Inferência estatística versus inferência causal

First, I would like to commend the *Revista Brasileira de Saúde Ocupacional* (RBSO) and Rita Fernades, Veronica Lima e Fernando Carvalho, the authors of "The critical use of statistical inference in occupational epidemiology: essay"[1] for bringing up the very important issue of the (mis)use of statistics in Epidemiology. Although it is old and has already been discussed and documented, its recurrence in the literature is still welcome, especially in Portuguese. The regrettable fact that it is still not even properly recognized by many as a serious "problem" in the development of scientific knowledge in general and, more particularly, in our field, reinforces this need. So, in keeping with its content, relevance, and appropriateness, the article is another text that deserves applause and our careful consideration.

Categorically, I dare say that the central point for moving forward in this discussion is to recognize the important and perhaps misunderstood distinction between "statistical inference" and "causal inference". And in addition to studies that measure the severity of a certain health problem, it is precisely with the aim of making causal inferences that epidemiological study designs, whether experimental or observational, are concentrated. Therefore, in an attempt to justify and defend this position, I will present some comments, counterpoints, and reflections similar to the issues raised in the article, which have also long been present in my concerns and writings[b]. I will explore more general issues, which I believe are little perceived as fundamental to this debate. I will leave aside more specific issues, such as the use of confidence intervals to consider a result statistically significant or not, which is addressed in the article and well explored in the literature[3].

Although the article explicitly touches on this aspect of causality, its main concern is to draw attention to the mandatory presence of randomness in the sampling process if statistical inferences are to be made and, in a somewhat questionable way, even causal inferences. In addition, with the aim of delimiting the uses of statistics in epidemiological studies and perhaps minimizing misunderstandings of these uses, the article proposes that studies be classified according to their purpose as descriptive, analytical, or inferential. With all due respect and acknowledging its appropriateness, I don't believe that this classification contributes significantly when we want to use statistical theory and practice with the explicit intention of making causal inferences. Furthermore, while it is

[b] Particularly in several chapters of the book entitled "Epidemiologia", published by Editora Atheneu, the third edition of which is due to be released at the 12th Brazilian Congress of Epidemiology, in Rio de Janeiro, at the end of November 2024. another reference I would highlight is the 2002 book "Inferência causal em epidemiologia"[2].

common ground that the random selection of the sample is a *sine qua non* condition for making statistical inferences - this being exactly one of the central points of misunderstanding -, it is controversial when the focus is on establishing causality.

In causal inference, the central issue is internal validity. In other words, using the consolidated classification of the literature[c], the concern is with the classic selection and information biases and with the complex and challenging phenomenon known as "confounding". This, in a nutshell, refers to a lack of comparability that occurs in the population. In other words, conceptually speaking, confounding is not a bias in the epidemiological sense, it is not an **error** made by the researcher, but it does bias the estimate of a causal effect, in a statistical sense, therefore. In this sense, in a cohort study, for example - or in any other observational study whose intention is to identify risk factors, in other words, to unveil causal links - even if we select a random sample from the exposed population and, also, from the unexposed population, these groups may still not be comparable due to some confounder. And, unless we can control the association by this confounder (if we already suspected it, and it was measured!), the estimate of the causal effect will be biased. The same goes for cross-sectional studies[d].

Randomization, that is, the process of randomly allocating a certain supposed causal agent, if possible, is the statistical mechanism that, in theory, or rather, on average, would probabilistically "guarantee" a causal inference (provided, of course, that there are no other alternative explanations, other biases), since it has the "power" to control known and even unknown potential confounding variables. As for whether, or not, there was a random selection of the units that were randomized, in other words, whether external validation makes sense, the generalization of this result to a certain population, in this inferential sense, is a secondary issue. Given the inherent interpersonal variability, even potentially referring us to the concept of interaction, once it is recognized and accepted (never demonstrated) that a certain condition may have a causal effect - at least on average, which is the "maximum" we can do from a well-planned study - it is the biological, social, cultural etc. profile of the people in that population that would "validate", or not, any causal extrapolations, in other words, outside the scope of statistical inference. The issue of sample representativeness is therefore not essential in causal inferences. For example, suppose that a well-designed randomized clinical trial conducted only on a non-random sample of white male workers concluded that there was a significant (statistically and/or practically relevant) causal effect of certain personal protective equipment. Why couldn't we readily generalize this result to non-white men or even women?

In short, I think that the problem seems to occur when we are explicitly interested in making causal inferences and believe that all the powerful theoretical and practical support of statistics, including descriptive statistics, would eventually be sufficient for such a challenge. In fact, it is not, but there is a widespread perception that it is. This perception, however, although mistaken, may have its "explanations". Firstly, the absolute lack of methodological and theoretical training in the field of causal inference, which is not only fairly recent, but can also be quite complex. On the other hand, statistics has revolutionized scientific knowledge to such an extent since the last century[5] that it has perhaps taken root in the collective imagination of academia as the cornerstone of scientific reasoning, especially because of the omnipresent and omnipotent sentence "this result is statistically significant". Since, in fact, statistical procedures are necessary in any analysis of empirical data

[c] This classification, although well-established, may not be complete precisely because of statistical reasons. Greenland (1991)[4] proposed a much more interesting classification based on the concept of validity, but unfortunately it didn't "catch on" in the literature, perhaps because it was published as a chapter in a not very accessible compendium, despite being in its 7th edition. In this classification, we can recognize confounding as "Comparison validity", selection bias as "Follow-up validity", and information bias as "Measurement validity". What is new is what he called "Specification validity", which relates to all the possible assumptions of the statistical analyses used in the causal inference process.

[d] A parallel reflection is in order here, but it is also related to these misunderstandings associated with statistics in epidemiology. Some, perhaps many, are radical in stating that cross-sectional studies do not lend themselves to causal inferences basically because it cannot be immediately guaranteed that the exposure in question precedes the outcome temporally. Although, in fact, in this type of study, the potential problem known as "reverse causality" cannot be ruled out by the available data, this is not usually the most serious problem. Depending on what the exposure and disease are, knowing a little about their pathophysiology, possible reverse causality could be ruled out. For example, when studying the causal association between smoking and a certain type of cancer, it doesn't make much sense to think about reverse causality. What does weaken a cross-sectional study, but is not a categorical impediment, is the potential prevalence bias (or survival bias). In other words, when we study only the prevalent cases of a certain disease, and not the incidents, even if they are all the existing ones, we are only looking at a "sample" (not in a statistical sense) which may not be representative of all the cases.

from a certain population - whether they are sample-based or not, whether they are derived from well-planned study designs or not, whether they are randomized or not - it seems that this appeal has gained "believers" throughout the academic world, particularly in the biomedical field[6]. Thus, one might suspect that the difficulties intrinsic to the process of causal inference have led to an overvaluation of any observed associations, simply because they are statistically significant. The p-value, i.e. the criterion of statistical significance (or even a confidence interval) refers only to random **error** and is therefore absolutely insufficient for establishing cause and effect relationships.

In this context, the use and interpretation of the p-value in biomedical research has been a point of intense debate and controversy for over 30 years, even forcing the American Statistical Association (ASA) to issue an official statement[7]. It's not that there's anything wrong with the theory behind statistical hypothesis testing, which results in the level of significance, but there have systematically been serious misunderstandings in its interpretation, which has compromised scientific development, as many have denounced in dozens of scientific articles and even books. For example, the interpretation, with its consequences, of what is or is not statistically significant from a previously fixed cut-off point (the fateful $p < 0.05$), has been criticized[8], although there are defenders and, recently, even the proposal of a stricter significance level, to 0.005[9,10].

To flesh out the ideas, let's go a little further in this discussion, but still with conceptual reflections based on a classic and emblematic example cited in the article: how can we affirm, with scientific "certainty", that it would be **true** that smoking **causes** lung cancer, as literature has been accumulating evidence for many years? To put it in a statistical-epidemiological way, how much more likely would the incidence of lung cancer be in smokers compared to non-smokers? Are we in a position to make this kind of inference validly and accurately? The emphasis on the words "true" and "cause" in this paragraph and on "error" in two previous moments was deliberate. They deserve further reflection, and we can connect them.

First, we must recognize that the concept of cause itself is very complex, with roots in philosophy. To simplify and strategically "escape" from this conceptual imbroglio, we can understand a cause, in this context, as any condition that alters the risk (i.e. the probability, the incidence) of a certain disease occurring. And, even more strategically, in order to make the insertion of statistics feasible, we can focus on defining and estimating causal parameters, such as relative risk, for example. That's it, this first stage seems well resolved.

Now, establishing, knowing or recognizing a truth is no simple task, if not impossible, as we can see from philosophy. But some ideas help. For example, Hanna Arendt is credited with the idea that we can think of three types of truth: factual, philosophical, and scientific. It's easy to characterize them, or understand them, by their opposite meanings. The opposite of factual truth is a lie, that of philosophical truth is illusion, and, tellingly, the opposite of scientific truth is **error**. This idea is very pertinent and appropriate to epidemiological research. Thus, from the scientific evidence accumulated so far, it is extremely unlikely that the established truth that smoking causes lung cancer is wrong. So why is it so difficult to establish an epidemiological "truth"? Precisely because of the omnipresence of **error** in research, whether of a random nature or, above all, due to some kind of bias, in a broad sense.

Another point that I identify as fundamental to be more attentive to, although perhaps well recognized as a hypothesis, is that statistical theory doesn't exactly lend itself to "understanding" singularities. However, in any statistical analysis of data there will always be a certain implicit assumption - untestable - of the homogeneity of the units. In this way, perhaps unlike other areas of application, the interpretation of statistics from research in the biomedical field has a peculiar character, and a fundamental contribution to understanding this process can come from Epidemiology, aided and complemented by Biology[11]. Consequently, in agreement with the article[1], the results of a particular study should contribute more to increasing knowledge about a certain phenomenon than to "revolutionizing" the theory about it. The subtle issue here is that, using statistical jargon, study observations are commonly assumed to be realizations of "i.i.d." random variables, i.e. independent and identically distributed. Whether they are independent or not is easier to recognize and we have analytical strategies to deal with both situations. The "dangerous" detail lies in the "identically distributed" premise! And here, without further ado, is a provocative question for reflection: should the statistical results of a survival analysis in a well-conducted study with 100 cancer patients be as scientifically "enlightening" as another equally well-conducted and well-analyzed study with, for example, 100 light bulbs?

In summary, despite the dozens of formal definitions of statistics and biostatistics, informally we can say that statistics is the science of uncertainty and variability, and biostatistics is this science applied to the field of health, considering two important peculiarities of this field: the intrinsic heterogeneity of the units of analysis (people) and the central interest in establishing cause and effect relationships. I believe that these reflections can help us to better interpret the statistical results of an epidemiological or clinical study.

References

1. Fernandes RCP, Lima VMC, Carvalho FM. The critical use of statistical inference in occupational epidemiology: essay. Res Bras Saude Ocup. 2024;49:e13. Available from: https://doi.org/10.1590/2317-6369/35622en2024v49e13
2. Luiz RR, Struchiner CJ. Inferência causal em epidemiologia: o modelo de respostas potenciais. Rio de Janeiro: Editora Fiocruz; 2002.
3. Gardner MJ, Altman DG. Confidence intervals rather than p-values: estimation rather then hypothesis testing. British Medical Journal, 292:746-750, 1986. https://doi.org/10.1136/bmj.292.6522.746
4. Greenland S. Concepts of validity in epidemiological research. In: Holland WW, Detels R, Knox G, editors. 2nd ed. New York, Oxford University Press; 1991. (Oxford textbook of public health, vol. 2).
5. Salsburg DS. Uma senhora toma chá…: como a estatística revolucionou a ciência no século XX. Rio de Janeiro: Zahar; 2009.
6. Salsburg DS. The religion of statistics as practiced in medical journals. Am Statistician, 1985;39(3):220-3. https://doi.org/10.2307/2683942
7. Wasserstein RL, Lazar NA. The ASA's Statement on p-values: context, process, and purpose. Am Statistician.2016;70(2):129-33. https://doi.org/10.1080/00031305.2016.1154108
8. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SNet al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 2016;31:337-50. https://doi.org/10.1007%2Fs10654-016-0149-3
9. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R. et al. Redefine statistical significance. Nature Human Behaviour. 2018; Jan;2(1):6-10. https://doi.org/10.1038/s41562-017-0189-z
10. Ioannidis JPA. The proposal to lower p-value thresholds to .005. JAMA. 2018;319(14):1429-30. https://doi.org/10.1001/jama.2018.1536
11. Schlesselman JJ. Biostatistics in epidemiology: a view from the faultline. J Clin Epidemiol. 1996 June;49(6):627-9. https://doi.org/10.1016/0895-4356(96)00036-4

Author contributions: Luiz RR conceived the argument, collected the information, drafted the Discussion and approved the revised and published version. The author takes full responsibility for the content published.

Data availability: The entire data set supporting the results of this discussion has been published in the article itself.

Funding: The author declares that his work in this Discussion was not funded.

Competing interests: The author declares that there are no conflicts of interest.

Presentation at a scientific event: The author informs that this Discussion is original and was not presented at a scientific event.

Received: August 12, 2024
Approved: August 13, 2024

Editor-in-Chief:
Eduardo Algranti

Ronir Raggio Luiz

Inferência estatística versus inferência causal
Statistical inference versus causal inference

Revista Brasileira de Saúde Ocupacional
vol. 50, edisfl1, 2025
Fundação Jorge Duprat Figueiredo de Segurança e Medicina do Trabalho - Fundacentro,
ISSN: 0303-7657
ISSN-E: 2317-6369

DOI: https://doi.org/10.1590/2317-6369/00124pt2025v50edisfl1