

André Luis Santiago Maia^a <https://orcid.org/0000-0002-8646-1486>

^aFundação Jorge Duprat Figueiredo de Segurança e Medicina do Trabalho, Escritório Avançado na Bahia. Salvador, BA, Brazil.

Contact:

André Luis Santiago Maia

E-mail:

andre.maia@fundacentro.gov.br

Hot to Cite (Vancouver):

Maia ALS. Common statistical and methodological errors in rejected manuscripts: the case of RBSO. Rev Bras Saude Ocup [Internet]. 2025;50:e11. Available from: <https://doi.org/10.1590/2317-6369/nt125en2025v50e11>



Common statistical and methodological errors in rejected manuscripts: the case of RBSO

Erros estatísticos e metodológicos comuns em manuscritos rejeitados: o caso da RBSO

Abstract

The Revista Brasileira de Saúde Ocupacional (RBSO) receives several manuscripts with studies involving quantitative methods in Occupational Safety and Health (OSH). However, many submissions present methodological and statistical errors that compromise their acceptance. Lack of clarity in results and inadequate use of statistical methods are recurring issues. Insufficient description of methodologies, poorly justified statistical analyses, and unwarranted extrapolation of results are also frequent. This editorial note discusses common errors observed in manuscripts rejected by RBSO, with the aim of improving the quality of publications in the OHS field.

Keywords: Statistics as Topic; Investigative Techniques; Occupational Health.

Resumo

A Revista Brasileira de Saúde Ocupacional (RBSO) recebe uma série de manuscritos que relatam estudos que envolvem métodos quantitativos em Saúde e Segurança do Trabalhador (SST). Entretanto, muitos textos apresentam erros metodológicos e estatísticos, que comprometem sua aceitação. Falta de clareza nos resultados e uso inadequado de métodos estatísticos são falhas recorrentes. A descrição insuficiente das metodologias, análises estatísticas mal justificadas e a extrapolação indevida de resultados também são frequentes. Esta nota editorial discute erros comuns observados em manuscritos rejeitados pela RBSO, com o intuito de elevar a qualidade de publicações na área de SST.

Palavras-chave: Estatística como Assunto; Técnicas de Pesquisa; Saúde do Trabalhador.

Every year, the Revista Brasileira de Saúde Ocupacional (RBSO) receives several manuscripts with studies that reflect the diversity and complexity of research into Occupational Safety and Health (OSH). When statistical methodologies are involved, the texts undergo preliminary evaluation by the journal's statistics department. At this stage, careful analysis reveals recurring limitations that often compromise the quality of potential publications. When manuscripts do not meet the criteria required in this initial evaluation stage, they are rejected.

Among the main challenges in evaluating manuscripts submitted to RBSO is the lack of clarity and coherence in the objectives, methods, and quantitative results. Texts with a disjointed structure and poorly organized ideas hinder the reviewer's understanding. The presentation of information should follow a logical sequence that enhances comprehension, with well-structured sections and paragraphs.

A superficial literature review, with significant gaps in the contextualization of the study and a lack of critical analysis of previous research, often results in a weak theoretical foundation. The literature should be reviewed exhaustively and critically, placing the study in the context of the research and highlighting its original contribution.

The statistical review pays particular attention to errors in study design, data analysis, interpretation, and presentation of results.

The description of the methods used is a critical point. When authors fail to detail the data collection procedures, doubts are raised about the validity of the results. It is essential that the methods are well described, allowing the study to be replicated by other researchers and guaranteeing the robustness of the findings.

To improve the reporting of studies, according to the instructions to RBSO authors, it is recommended to use scientific writing guides. A valuable source is the EQUATOR network portal - Enhancing the Quality and Transparency of Health Research (<https://www.equator-network.org/>), where there is a comprehensive library that provides access to publications related to writing guides. Other materials of interest are also available, for example on publication ethics, and guidance materials for editors, reviewers, and authors.

Drawing up and implementing an analysis plan is important to avoid presenting statistical analyses without proper justification or using methods that are not the most appropriate for the type of data or hypotheses in question. Insufficient or inadequate data analysis compromises the interpretation of the results. A consistent analysis combines thorough selection of statistical methods with careful interpretation of the results.

A common mistake is to draw conclusions that are not supported by the quantitative results presented, i.e. to extrapolate the findings, suggesting implications that are not supported by the data. Conclusions should be based directly on the results, avoiding undue generalizations and staying within the limits of what was actually found.

Confusion between the significance level (α) and the p-value is another common misconception. Although they are related, they have different concepts in statistics. Many manuscripts present these measures incorrectly, without differentiating between them. α is a pre-established value that represents the maximum probability of rejecting the null hypothesis (H_0) when it is, in fact, true¹. The p-value, in turn, is obtained from the test statistics and represents the probability of obtaining results as extreme (or more extreme) than those observed, assuming that H_0 is true². The lower the p-value, the lower the probability that the observed results occurred by chance.

There is also a tendency to conclude on associations between variables based solely on linear correlation coefficients, without considering other measures that could provide a more accurate understanding. Many relationships between variables are non-linear, and the linear correlation coefficient does not capture these nuances. A scatter plot can reveal non-linear patterns that correlation coefficients do not identify.

The Chi-square test is based on a fundamental premise regarding expected frequencies³. When the expected frequencies in a contingency table are too low, the Chi-square distribution may not be accurate, compromising the validity of the results and leading to false conclusions about the association between the variables. Applying the test without respecting this premise is another common flaw in manuscripts submitted to RBSO. It is advisable to explore alternative tests that are better suited to the specific characteristics of the data.

When the study involves hypothesis testing, it is essential to describe in advance the hypotheses being tested, as well as clarifying the statistical method applied and the assumptions for the validity of the test. This is not just a matter of organization, but an essential practice to ensure that the results are correctly understood by the readers.

The use of non-probability samples in research with subsequent application of statistical inference techniques is the main problem in the studies reported in the manuscripts. The distinction between probability and non-probability samples is central to statistics. Probabilistic samples are selected so that each unit of the population has a known, non-zero chance of being included in the sample⁴. This makes it possible to generalize the results obtained to the entire target population, which corresponds to the group the researchers are trying to understand⁵. Although most studies are conducted using sample data, the target population must be identified and described in the manuscript. This information is essential for the reviewer to be able to judge the external validity of the study.

In methods, it is sometimes described that an effort has been made to calculate the sample size based on probability sampling theory, but the principle of randomness is ignored, and the final sample is made up of volunteers. This is a classic example of convenience sampling, which can introduce significant biases, such as selection bias. This type of bias can occur when the individuals willing to take part in the study have different characteristics from the target population, which can distort the results⁶. Porta et al.⁷ give an example of the estimated effect of smoking on heart disease. According to the authors, the estimates will be biased if the study participants are volunteers and the decision to take part is influenced by smoking and a family history of heart disease.

It is essential to recognize the limitations of convenience samples, avoiding hasty generalizations for the target population. The principle of randomness is fundamental to the validity of inferential statistical techniques based on the normal distribution, such as estimating population parameters and constructing confidence intervals⁸. Estimates from non-probabilistic samples can be biased and may not accurately reflect population parameters. When analyzing non-random samples, statistical bias correction techniques can be explored to minimize some of the associated problems⁹.

It is important to clearly describe the sampling procedure, i.e. how the participants included were recruited and selected, whether in a probabilistic or non-probabilistic sample. This information is crucial for assessing the relevance of the methods used, as well as for analyzing potential biases in the study.

It should be noted that probability samples are not always necessary or feasible. There are various study designs that make it possible to investigate the occurrence of diseases and injuries related to occupational exposures. The choice of study design is guided by factors such as: the nature of the outcome of interest, its relationship with the exposures studied, the frequency of the outcomes and exposures, the characteristics of the target population, the data needed to carry out the study, its availability, or the feasibility of collecting it¹⁰.

In results, before the reader comes across a table or graph, it is crucial that they understand what will be presented and why this element is relevant. The prior description provides the necessary context to correctly interpret the statistics that will be shown and how they relate to the study's hypotheses or conclusions. For these reasons, the detailed description must not be omitted from the manuscript.

Many authors insist on using the symbol " \pm " to represent the standard deviation (SD). This wrongly suggests that there is a statistical range "mean - SD" to "mean + SD", which does not reflect the true meaning of SD. The appropriate way is to present the SD in brackets preceded by the mean. This avoids misinterpretation.

Many manuscripts fail to provide details about the regression models used. The justification for choosing the specific model and information on how it was constructed should be described, as well as the statistical assumptions of the model. The results for the statistics used to assess the quality of the fit should be presented.

The conclusions drawn from the fitted models are sometimes superficial or incorrect, without a critical interpretation of the coefficients or without considering the limitations of the method. These shortcomings compromise the validity of the results obtained from regression analysis and prevent a proper statistical evaluation.

Hosmer et al.¹¹ is a complete reference for the specific case of logistic regression. It covers the construction, interpretation, and evaluation of models, presenting the fundamentals and more complex applications.

Finally, the lack of methodological review and of clarity in the presentation of results is evident in several quantitative analysis texts received by RBSO. Statistical techniques must be understood to assess the validity and clarity of the results obtained. The methodological review must be treated rigorously to ensure that the text is free of errors or gaps that could jeopardize its evaluation.

It is hoped that the authors will be able to use these reflections to improve their future work, strengthening their methodological base. It is believed that this initiative will encourage the scientific community to increasingly seek statistical rigor in their investigations and will contribute to raising the quality of manuscripts reporting quantitative research in OSH, especially in RBSO.

References

1. Neyman J, Pearson ES. *On the problem of the most efficient tests of statistical hypotheses*. Philosophical Transactions of the Royal Society A. 1933 Feb.;231:289-337. <https://doi.org/10.1098/rsta.1933.0009>
2. Fisher RA. *Statistical methods for research workers*. Edinburgh: Oliver and Boyd; 1925.
3. Snedecor GW, Cochran WG. *Statistical methods*. Hoboken: Wiley-Blackwell; 1989.
4. Cochran WG. *Sampling techniques*. Hoboken: John Wiley & Sons; 1977.
5. Mohr DL, Wilson WJ, Freund RJ. *Statistical methods*. Cambridge: Academic Press; 2021.
6. Pearce N, Checkoway H, Kriebel D. Bias in occupational epidemiology studies. *Occup Environ Med*. 2007;64(8):562-68. <https://doi.org/10.1136/oem.2006.026690>
7. Porta MS, Greenland S, Hernán M, dos Santos Silva I, Last JM, eds. *A dictionary of epidemiology*. 6th ed. Oxford: Oxford University Press; 2014. <https://doi.org/10.1093/acref/9780199976720.001.0001>
8. Casella G, Berger R. *Statistical inference*. Boca Raton: CRC Press; 2024.
9. Baker R, Brick JM, Bates NA, Battaglia, M, Couper, MP, Dever JA et al. (2013). Summary report of the AAPOR task force on non-probability sampling. *J Surv Stat Methodol*. 2013;1(2):90-143. <https://doi.org/10.1093/jssam/smt008>
10. Checkoway H, Pearce N, Kriebel D. Selecting appropriate study designs to address specific research questions in occupational epidemiology. *Occup Environ Med*. 2007 Sep;64(9):633-8. <http://doi.org/10.1136/oem.2006.029967>
11. Hosmer Jr DW, Lemeshow S, Sturdivant RX. *Applied logistic regression*. Hoboken: John Wiley & Sons; 2013.



Available in:

<https://www.redalyc.org/articulo.oa?id=100582247001>

How to cite

Complete issue

More information about this article

Journal's webpage in redalyc.org

Scientific Information System Redalyc
Diamond Open Access scientific journal network
Non-commercial open infrastructure owned by academia

André Luis Santiago Maia

**Erros estatísticos e metodológicos comuns em
manuscritos rejeitados: o caso da RBSO**

**Common statistical and methodological errors in rejected
manuscripts: the case of RBSO**

Revista Brasileira de Saúde Ocupacional

vol. 50, e11, 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/nt125pt2025v50e11>