EUROPEAN JOURNAL OF BIOMEDICAL AND PHARMACEUTICAL SCIENCES
http://www.ejbps.com

ISSN 2349-8870
Volume: 11
Issue: 10
13-18
Year: 2024

THE MOST COMMON STATISTICAL ERRORS FOUND IN ALBANIAN MEDICAL RESEARCH PUBLICATIONS: CRITICAL INSIGHTS FROM A RANDOM STUDY SAMPLE OF FREE ONLINE PUBLISHED ARTICLES AND LEARNINGS FROM THE LITERATURE

Dr. Leidon Shapo[1]* and Eduart Hashorva[2]

[1]MD, Public Health Lead (Adults), Berkshire, UK.
[2]Assoc. Prof., MD, PhD, Obstetrics & Gynecology University Hospital "Geraldine Queen", & Catholic University "Lady of Good Council" Tirana, Albania.

***Corresponding Author: Dr. Leidon Shapo**

MD, Public Health Lead (Adults), Berkshire, UK.

Article Received on 05/08/2024 **Article Revised on 25/08/2024** **Article Accepted on 15/09/2024**

SUMMARY

Background: There is now strong evidence to support the statistical analysis importance as part of any medical research. However, the study method and the inappropriate use of statistical analysis can lead to study invalidity and false conclusions. **Study Method:** This is a review article aimed to highlight some of the most common medical statistical errors found from an internal observation of published articles in three of the Albanian medical journals (available free online) during a five-year period (2019 – 2023). In this paper we present the results from a sample of fifty randomly selected articles that met the eligibility criteria. The evaluation process included the statistical method used with a focus on identifying the most common errors present in these studies while we reflect on learnings from the literature on the same subject and specific statistical errors. **Results:** The top three most common statistical errors we found were: the 'lack of the sample size calculation (88%)', the 'sampling selection criteria not provided (64%)', and the 'absence of an adequate control condition/group (50%)'. **Conclusions:** The statistical methods used in any medical research are key components as they provide both the required study accuracy and reliability. In addition, the use of appropriate statistical methods supports the process of making correct statistical analyses, draw the right conclusions and avoid various errors. The misuse of statistics in medical studies is a known issue that is both unethical and can have serious clinical consequences.

KEYWORDS: Medical research, study design, statistical methods, statistical errors, confidence interval.

INTRODUCTION

There is an old saying that: "*Statistics is like a cart that takes us where we need to go*". Today, the statistical analysis is a very important part of any research study in medicine. But the inappropriate use of statistical analysis can lead to false conclusions. Unfortunately, some statistical mistakes are common, which can falsely legitimize data. The misuse of statistics in medical studies has been treated widely in the literature and the agreed conclusion is that it is both unethical and can have serious clinical and public health consequences. There are several data errors that can contribute to incorrect conclusions, and as consequence can compromise the validity of studies, while overestimating or underestimating the effects of a specific treatment.

To avoid making these errors, it is critical to assess them and better understand any statistical concept used. This process is expected to lead to the use of appropriate statistical techniques for specific research questions and the calculation of an appropriate sample size to guarantee adequate statistical power.

Some of most common statistical errors in medical research include sampling bias, the incorrect determination of the study sample, failing to adjust for multiple comparisons, misinterpreting p-values as a measure of effect size or clinical relevance, choosing incorrect tests for a particular data set, type I and II errors, data fishing, and publication bias. It is also important that researchers interpret their results using appropriate statistical concepts by soliciting feedback from specialist statisticians and reflecting any gaps as part of the study limitation section.

There is evidence to suggest that the above mistakes are often interdependent, such that one mistake will likely impact others, which means that many of them cannot be

remedied in isolation. Moreover, there is usually more than one way to solve each of these mistakes: for example, we focus on frequentist parametric statistics in our solutions, but there are often Bayesian solutions that we do not discuss.

STUDY METHOD

This is a review study aiming to both present some of the most common statistical errors found in a random sample of online published articles (Albanian journals), and critically appraise the findings on the lights of available evidence from the literature.

In order to do so, we have (i) randomly selected a total of fifty research articles published online in medical journals of Albania over a period of five years (2019 - 2023).

Sampling: The selection was made using the free online archives of the three most important Albanian medical journals. Using a probability sampling technique that involved a selection of a subset of online articles, where each published online paper had an equal chance of being selected; we randomly and retrospectively selected every third article that met the criteria to have statistical processing, until we reached 50 cases, which made up 20% of the total number of articles published in these journals for the 5-year period. Statistically, the final number was sufficient for the purpose of this review. Furthermore, we created a database in Excel with the errors found in these 50 articles, which we have analyzed and presented the findings in this study.

We must emphasize the importance of using the findings from this sample as an illustration of most common statistical errors described and discussed on the literature and that's where the real value of this study lies.

Inclusion criteria: We used a simple selection technique that randomly generated fifty articles using the study eligibility criteria that only included articles reporting statistical findings and analyses, *excluding* any studies using qualitative research methods such as depth interviews, diary studies, participant observations, editorials, mini-reviews or conversational analyses that did not include the use of statistics.

The final sample of fifty articles was then (ii) analyzed to assess the statistical method used, and the key elements that could lead to most common errors presented in all these studies as defined in the medical research literature. All the key findings are presented in Tab. 1.

The data were treated according to the code of ethics, preserving the anonymity of the authors and journals.

Study limitations: We are aware that what is presented as key findings in this review in terms of study sample (online studies published in Albanian medical journals) is probably not sufficiently generalizable to other medical journals. However, in our discussion section below, we consider the most common statistical errors based on the findings from our random sample, but at the same time also provide considerations and reflect on the evidence from the literature for each error listed. In our opinion that is the external value that could apply to a much larger medical research audience.

RESULTS

The results of our study are summarized in Tab. 1.

Tab. 1: Top 10 statistical Errors of Statistics in Medical Research.

Nr.	Most common statistical errors	Frequency (50)	
		Number	%
1	Lack of Sample size calculation	44	88
2	Sampling Selection criteria no given	32	64
3	Absence of an adequate control group	25	50
4	Wrong interpretation of Correlation and causation	24	48
5	Inappropriate use of parametric test	22	44
6	Overall inappropriate interpretation	21	56
7	Use of wrong statistical analysis	18	36
8	Statistical technique defined but not used	16	32
9	Statistical technique used but not defined	14	28
10	Errors in interpreting probability (p) values	12	24

DISCUSSION

In this review, we focus on the most common statistical errors found in the final subset of articles published online in Albanian medical journals and the value of what we can learn from each error as described in the literature.

1. Sampling error

This is also known as the minimum required number included in any medical research study and is the most frequent statistical error.

We found this type of error present in as many as 44 (88%) of our randomly selected and online published papers.

Learnings from the literature: A small sample size does not mean that the results are wrong but rather that the data is consistent with a wide range of possible hypotheses. A wide interval for example means that the sample size was too small and cannot provide any meaningful information about the value of a treatment.

Power analyses: Typically, the smaller the sample size, the larger any difference between the study groups compared and their scores will have been in order to achieve statistical significance. Statistical power analysis is a set of procedures and formulas that allow us to determine how likely we would achieve statistical significance with a particular sample size given that there is a true difference between groups. If the likelihood is good (i.e. greater than or equal to an 80% chance), then the sample size is considered adequate. A power analysis can be used to calculate the minimum sample size required so that one can be reasonably certain to detect an effect. The power of a statistical test is the probability that the test will reject the null hypothesis when the alternative hypothesis is true.

Increasing the sample size involves tangible costs, both in time, money, and effort therefore it is important to make the sample size large enough but not wastefully large.

The population variability also affects the sampling error. More variable populations give rise to larger errors as the samples, or the estimates calculated from different samples are more likely to have greater variation. The effect of the variability within the population can be reduced by increasing the sample size to make it more representative of the survey population.

1. Sampling Selection criteria

The first step in developing a full study protocol is having a clear understanding of the research question to be answered. The next step is to have explicit definitions of the independent and dependent variables of interest in the study.

Sample selection criteria are the set of characteristics that must be present when selecting participants for a study. Sample selection criteria often include age ranges, diagnoses, gender, and other characteristics specific to the research question.

We found this type of error in 32 (64%) of articles included in our sample of studies published online.

Learnings from the literature: There are some basic elements related to the selection of participants for health research. Sample representativeness, sample frame, types of sampling, as well as the impact that non-respondents may have on results of a study are described.

There are several theoretical and practical reasons that prevent us from carrying out census-based surveys, including: ethical issues, budgetary limitations, logistics, time restrictions, and unknown target population size. All these reasons explain why samples are more frequently used. However, researchers must be aware that sample results can be affected by the random error (or sampling error).

The sample frame – that's the group of individuals that can be selected from the target population given the sampling process is going to use in the study. For example, to identify cases of a particular skin cancer the researcher may consider utilizing as sample frame the national cancer registry system or the anatomopathological records of skin biopsies. Given that the sample may represent only a portion of the target population, the researcher needs to examine carefully whether the selected sample frame fits the study objectives or hypotheses, and especially if there are strategies to overcome the sample frame limitations.

2. Failure to use an adequate control group

The purpose of a control group is to allow the observer to conclude that any change observed in the "active treatment group" is due to the treatment being studied, rather than to other factors.

We found the lack of a proper use of a control group in 25 (50%) of our randomly selected articles.

Learnings from the literature: Control groups are particularly important when factors in addition to the intervention under study can affect the outcome of interest, when the new technology or treatment of interest and some established technology or existing therapy are both effective, and when the natural course of untreated disease is not clear or consistent, as is the case with breast cancer. Failure to use a control group, or use of an inappropriate control group, can make it impossible to draw meaningful conclusions from a study.

Given the purpose of a control group, it is important that patients in the treatment and control groups be similar in terms of baseline characteristics that can influence the outcome of the intervention under study. For example, if one study group included more women at high risk for cervical cancer than another group, then a detection technology tested in the high-risk group would likely detect more cancer cases than a technology tested in the low-risk group, leading to the perception that the detection system was more sensitive than a system tested in lower risk patients.

3. Interpretation of Correlation and causation

Causation, by definition, implies that one event is the result of another event occurring. Correlation is a statistical measure of the (linear) relationship between two variables. If one has data on any two variables, one may compute the correlation between them.

A correlation between variables, however, does not automatically mean that the change in one variable is the cause of the change in the values of the other variable.

We found this type of error in 24 (48%) of our reviewed articles.

Learnings from the literature: Establishing a causal relationship between two variables is known to be quite challenging. An example of how to establish causation may be taken from the field of medical research. Most individuals have heard of the placebo effect. This process has historically been used in testing new medication. A group of individuals with nearly identical characteristics (and/or symptoms) is split into two (or more) groups. One group is given the new medication and one is given a placebo. A study is conducted to determine the treatment effectiveness. If the medication is successful and the placebo effect is negligible, one may begin to assume a causal relationship exists between the new medication and the effects of the disease or illness in the group.

One area some researchers often make mistakes when interpreting correlation is that correlation implies a linear relationship between the variables of interest. Even if you find that two variables could show some linear relationship relying only on computing a correlation coefficient may give someone a false read on the true relationship between those variables.

Strong correlation may be misleading and keep one from asking more in-depth questions. Immediately seeing strong correlation between two variables may lead one to work under the assumption of a linear relationship when a deeper construct lies below the surface waiting to be released.

The takeaway here is that if someone says that the correlation between two variables is statistically significant this should be considered only as the first step in determining causality.

4. Statistical tests: Inappropriate use of parametric test

In terms of selecting a statistical test, the most important question is "what is the main study hypothesis?" In some cases, there is no hypothesis; the investigator just wants to "see what is there". For example, in a prevalence study there is no hypothesis to test, and the size of the study is determined by how accurately the investigator wants to determine the prevalence. If there is no hypothesis, then there is no statistical test. The main disadvantage of parametric tests is that they are sensitive to violations of the assumptions, such as normality, homoscedasticity, or independence. If your data do not meet these assumptions, your results may be inaccurate or misleading.

We found this type of error in 22 (44%) of the randomly selected articles considered for this review.

5. Inappropriate interpretations

A bit surprising, but *in total, only 21 or 56% of the studies in our random sample did not provide complete and robust interpretation of statistical data and results*.

This could be due to lack of reference or electronic links to a scientific article; some form of misleading in terms of reporting – for example 'not reporting adverse events that were included in the scientific article', misleading interpretation such as claiming a causal effect despite non-randomized study design and over generalization/misleading extrapolation of the results such as extrapolating a beneficial effect from a specific small trial or an animal study to humans or larger groups. We also identified some cases such as highlighting a single patient experience for the success of a new treatment or intervention instead of focusing on the group results.

6. Use of wrong statistical analysis

It pretty much comes down to two things: whether the assumptions of the statistical method are being met and whether the analysis answers the research question. Assumptions are very important. A test needs to reflect the measurement scale of the variables, the study design, and issues in the data. A repeated measures study design requires a repeated measures analysis. A binary dependent variable requires a categorical analysis method.

We found this type of error in 18 (36%) of articles published online as part of our study pool.

7. Inappropriate use of p-values

P-values are often misinterpreted as a measure of effect size or clinical relevance, when in fact they only indicate the probability of observing the study results by chance. It is important to interpret p-values in the context of effect size and clinical relevance. For example, a study that finds a significant association between taking omega 3 and better control of type II diabetes with a p-value of 0.04, but the effect size is so small that it is unlikely to have clinical relevance. Based on this result, the researchers conclude that the treatment is effective and recommend it for clinical use. However, the p-value alone does not tell us the magnitude or clinical relevance of the treatment effect. It only tells us the probability of obtaining the observed difference in symptom reduction between the treatment and control groups if there is no true effect of the treatment. The effect size, on the other hand, tells us the magnitude of the treatment effect in terms of the size of the difference between the treatment and control groups. A small p-value can correspond to a large effect size, a small effect size, or something in between.

We found this type of error in 12 (24%) articles included on our final study sample.

8. Missing data

Although not included on Tab.1 and not reviewed as part of assessing our sample, 'Missing data' are a common occurrence in many study designs used to conduct primary care research (pragmatic randomized trials, observational studies, quality improvement efforts, etc.). The definition for missing data is when values are not available to the investigator that would have contributed to the final analysis had they been observed. Examples of missing data include patients lost to follow-up, partially filled-out surveys or incomplete medical records. Missing data can compromise the validity of study findings (e.g., risk for bias increases, subgroups of the population may be underrepresented, loss of information, reduced statistical power, etc.). Multiple studies have shown that high rates of missing data may negatively impact conclusions in primary care studies.

A holistic approach for addressing missing data should include approaches in the design phase (prior to the study), conduct phase (during the study) and in the analytic phase (after data collection is complete).

CONCLUSIONS

The statistical analysis are key component of medical research because it gives the study the required accuracy and reliability. The use of appropriate statistical methods in medical researches is a basic condition to make correct statistical analyses, to draw correct conclusions and to avoid possible mistakes.

In order to present biomedical research results with reliable and precise estimates based on data, it is extremely important to make researchers aware of the various aspects of statistics methods used in medical research as well as the role of engaging statistical experts need to be appreciated. Journals must have standard rules for reporting study design, sample size and data analysis tools. Careful attention is needed while applying any statistical procedure as, it will not only help readers to trust in the published articles, but also rely on the inferences the published articles draw.

In this perspective, we shall recommend to strengthen the eduction program of Medical statistics as part of the Albanian Faculty of Medicine curricula as well as the inclusion of epidemiology and medical statistics in the boards of various Albanian journals and any research team as a basic requirement.

REFERENCES

1. Altman DG. Statistics in medical journals: Some recent trends. *Statistics in Medicine*, 2000; 19: 3275-89.
2. Lang T, Secic M. How to report statistics in medicine: annotated guidelines for authors, editors, and reviewers. Philadelphia (PA): *American College of Physicians*, 1997.
3. Reed III JF, Salen P, Bagher P. Methodological and statistical techniques: what do residents really need to know about statistics? *Journal of Medical Systems*, 2003; 27: 233-8.
4. Huang W, LaBerge JM, Lu Y, et al. Research publications in vascular and interventional radiology: research topics, study designs, and statistical methods. *Journal of Vascular and Interventional Radiology*, 2002; 13: 247-55.
5. García – Berthou E, Alcaraz C. Incongruence between test statistics and P values in medical papers. *BMC Med Res Method*., 2004; 4: 13-7.
6. Porter AM. Misuse of correlation and regression in three medical journals. *J Roy Soc Med*., 1999; 92: 123-8.
7. Curran – Everett D, Benos J. Guidelines for reporting statistics in journals published by the American physio-logical association. *American Journal of Physiology and cell physiology*, 2004; 287: 243-5.
8. Gardenier JS, Resnik DB. The misuse of statistics: concepts, tools and a research agenda. *Account Res*., 2002; 9: 65-74.
9. Hanif A, Butt A, Awais SM. Importance and understanding of Biostatistics among post graduate students at King Edward Medical University Lahore, Pakistan. *ANNALS of KEMU*, 2009; 15 (3): 107-10.
10. Altman DG, Schulz KF, Moher D, et al. The CONSORT statement: revised recommendations for improving the quality of parallel – group randomized trials. *Annals of Internal Medicine*, 2001; 134: 657-62.
11. Altman DG, Goodman SN, Schroter S. How Statistical Expertise is used in Medical Research. *JAMA*, 2002; 287: 2817-20.
12. Strasak AM, Zaman Q, Pfeiffer KP, et al. Statistical errors in medical research – a review of common pit-falls. *Swiss medicine weekly*, 2007; 137: 44-9.
13. Altman DG. The scandal of poor medical research. *British Medical Journal*, 1994; 308: 283-4.
14. AbdulRaheem Y. Statistics in medical research: Common mistakes. *J Taibah Univ Med Sci*., 2023; 18(6): 1197-1199.
15. Makin TR & de Xivry OJJ. Science Forum: Ten common statistical mistakes to watch out for when writing or reviewing a manuscript. *eLife*, 2019; 8: e48175.17.
16. Kumar A, Kishun J, Singh U, Gaur D, Mishra P, Pandey CM. Use of appropriate statistical tools in biomedical research: Current trend & status. *Indian J Med Res.*, 2023; 157(4): 353-357.
17. Slutsky DJ. Statistical errors in clinical studies. *J Wrist Surg*., 2013; 2(4): 285-7.
18. Marino M, Lucas J, Latour E, Heintzman JD. Missing data in primary care research: importance, implications and approaches. *Fam Pract*., 2021; 38(2): 200-203.

19. Martínez-Mesa J, González-Chica DA, Duquia RP, Bonamigo RR, Bastos JL. Sampling: how to select participants in my research study? *An Bras Dermatol.*, 2016; 91(3): 326-30.
20. AbdulRaheem Y. Statistics in medical research: Common mistakes. *J Taibah Univ Med Sci*., 2023; 18(6): 1197-1199.