Journal of Taibah University Medical Sciences (2023) 18(6), 1197–1199

Taibah University

Journal of Taibah University Medical Sciences

www.sciencedirect.com

Letter to the Editor

Statistics in medical research: Common mistakes

Yousif AbdulRaheem

Professor in Community Medicine and Public Health, Board (equivalent to PhD) Degree, Fellow of Iraqi Scientific Council of Medical Specializations, Head of Medical Education Unit, Alkindy College of Medicine University of Baghdad, Iraq

Received 24 March 2023; accepted 12 April 2023; Available online 6 May 2023

Abstract

The misuse of statistics in medical studies has been discussed extensively with the conclusion that it is both unethical and can have serious clinical consequences. These errors can contribute to incorrect conclusions, compromise the validity of studies, and overestimate or underestimate the effects of treatment. To avoid making these errors, it is critical to consider their presence and understand statistical concepts. This practice will ultimately 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. Common statistical errors in medical research include sampling bias, the incorrect determination of 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 important that researchers interpret their results using appropriate statistical concepts by soliciting feedback from specialist statisticians.

Keywords: Error; Medical statistics; Research

E-mail: yousifabdulraheem@kmc.uobaghdad.edu.iq

Peer review under responsibility of Taibah University.

ELSEVIER Production and hosting by Elsevier

Dear Editor,

The statistical section is an important part of any research work because findings and inferences are not accurate unless they have been validated by various statistical tools and tests. Because most medical researchers lack significant statistical training and fail to work with someone who does, numerous errors are often made in data analysis and presentation in published medical studies.[1] If these errors are publicized, they tend to spread because new scholars notice them and attempt to replicate them. Fortunately, these errors are frequently identified during the peer-review process, and papers are returned for revision.[2] Here, we discuss some of the mistakes that can exert impact on the accuracy and reliability of research findings; it is important that medical professionals are aware of these issues.

1. **Sampling bias.** This occurs when the sample of study participants is not representative of the population being studied, and can lead to inaccurate conclusions relating to the effects of a particular intervention in that population or the inaccurate generalizability of the results. For example, a study on the prevalence of obesity that only includes participants from a particular geographic region; this will not be representative of the broader population. Another example is a study on the efficacy of a new treatment for a particular condition that only includes participants who have a milder form of the condition; this would overestimate the efficacy of the treatment in the general population.[3]
2. **Inappropriate sample-size calculation.** The validity of a study would be negatively affected if the calculated sample size was insufficient or incorrectly calculated. Samples should not be too large or too small, as both have limitations that can jeopardize research findings. A sample that is too small may prevent the results from being extrapolated, whereas a sample that is too large may amplify the detection of differences, thus emphasizing statistical differences that are not clinically significant. To

1658-3612 https://doi.org/10.1016/j.jtumed.2023.04.004

ensure adequate estimation of suitable sample size to detect a significant effect of the intervention and that the study has sufficient statistical power to detect such an effect, it is necessary to perform a sample size calculation.[4–6]

3. **Effect of confounding variables.** These are variables that are related to both the exposure and the outcome being studied and can lead to false associations. Controlling for confounding variables is important to isolate the true effect of the exposure. For example, a study on the relationship between meat consumption and heart disease that fails to account for other factors that may contribute to heart disease, such as smoking or physical activity, or a study on the effectiveness of a new treatment for tuberculosis that fails to account for differences in disease severity or other factors that may influence treatment outcomes.[7]
4. **Errors in the application of statistical tests.** Choosing the wrong test for a particular dataset affects the validation of observations during research. For example, the use of a parametric t-test instead of a non-parametric Wilcoxon signed-rank test when the data does not meet the assumptions of normality, or the failure to apply Yates' continuity correction to the chi-squared, particularly when the sample size analyzed is small.[8]
5. **Type I and Type II errors.** Type I errors occur when a study incorrectly concludes that there is a significant effect when there is not (erroneously finding a difference). These errors can result from the use of inappropriate statistical tests or inadequate sample sizes. For example, a clinical trial of a new drug concludes that it is effective based on statistical significance, but the effect size is very small and may not be clinically meaningful. This means that the study concludes that the drug is effective when it is not. Type II errors occur when a study incorrectly concludes that there is no significant effect when there is (the inability to identify a real difference). For example, a clinical trial of a new drug concludes that it is not effective based on statistical insignificance; however, in fact, the effect size is actually large and clinically meaningful. This means that the study concludes that the drug is not effective when it actually is.[9]
6. **Failure to adjust for multiple comparisons.** When multiple comparisons are performed in a study, the probability of a Type I error increases. To address this, statistical adjustments should be made to control the overall false positive rate. For example, a study is conducted to test the efficacy of a new drug for treating a particular condition. The study involves testing the drug on 100 patients and compares their outcomes to a control group of 100 patients who receive a placebo. The researchers measure several outcomes, including blood pressure, heart rate, and cholesterol levels. After analyzing the data, the researchers find that the drug appears to have a statistically significant effect on reducing blood pressure ($p = 0.03$), but not on heart rate ($p = 0.20$) or cholesterol levels ($p = 0.10$). Based on this finding, the researchers conclude that the drug is effective in reducing blood pressure and recommend it for clinical use. However, the researchers failed to adjust for multiple comparisons when analyzing the data. This means that they did not consider the fact that they tested multiple outcomes (blood pressure, heart rate, and cholesterol levels) and that there is a higher chance of obtaining a false positive result (i.e., a result that appears significant but is actually due to chance) when testing multiple outcomes.[3–10]
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. As reported by Lykken in 1968: "Statistical significance is perhaps the least important attribute of a good experiment; it is never a sufficient condition for claiming that a theory has been usefully corroborated, that a meaningful empirical fact has been established, or that an experimental report ought to be published."[8–11]
8. **Data dredging or "fishing".** This occurs when multiple analyses are conducted on a dataset to identify significant associations, without a clear hypothesis or theoretical framework. This can result in false positive findings that are not replicable. For example, a study that conducts multiple analyses on a dataset without a clear hypothesis or theoretical framework and reports a significant association between a particular variable and an outcome, even though the finding may be due to chance.[2–7]
9. **Publication bias.** This occurs when studies with positive or statistically significant results are more likely to be published than those with negative or non-significant results, thus leading to overestimation of the true effect size. Publication bias can occur for a variety of reasons, including the desire of researchers to publish positive results, the bias of journal editors and reviewers towards positive studies, and the tendency of researchers to focus on statistically significant results rather than the overall pattern of findings. Publication bias can have serious consequences as it can lead to incorrect conclusions about the effectiveness of treatments or interventions.[12]

These are just some of the common mistakes in medical research statistics. Overcoming these pitfalls, with carefully designed studies, and by interpreting the results in the context of clinical relevance and prior research objectives, will ensure the accuracy and reliability of study findings. This will eventually lead to better endeavor in medical research work.

Source of funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors have no conflict of interest to declare.

References

1. Bajwa SJ. Basics, common errors and essentials of statistical tools and techniques in anesthesiology research. **J Anaesthesiol Clin Pharmacol 2015 Oct- Dec**; 31(4): 547–553. https://doi.org/10.4103/0970-9185.169087. PMID: 26702217; PMCID: PMC4676249.
2. Altman DG. Statistics in medical journals: some recent trends. **Stat Med 2000 Dec 15**; 19(23): 3275–3289. https://doi.org/10.1002/1097-0258(20001215)19:23<3275::aid-sim626>3.0.co;2-m. PMID: 11113959.
3. Altman DG. Statistical reviewing for medical journals. **Stat Med 1998 Dec 15**; 17(23): 2661–2674. https://doi.org/10.1002/(sici)1097-0258(19981215)17:23<2661::aid-sim33>3.0.co;2-b. PMID: 9881413.
4. Devane D, Begley CM, Clarke M. How many do I need? Basic principles of sample size estimation. **J Adv Nurs 2004 Aug**; 47(3): 297–302. https://doi.org/10.1111/j.1365-2648.2004.03093.x. PMID: 15238124.
5. In J, Kang H, Kim JH, Kim TK, Ahn EJ, Lee DK, Lee S, Park JH. Tips for troublesome sample-size calculation. **Korean J Anesthesiol 2020 Apr**; 73(2): 114–120. https://doi.org/10.4097/kja.19497. Epub 2020 mar 16. Erratum in: Korean J Anesthesiol. 2020 Jun;73(3):268. PMID: 32229812; PMCID: PMC7113158.
6. Faber J, Fonseca LM. How sample size influences research outcomes. **Dental Press J Orthod 2014 Jul-Aug**; 19(4): 27–29. https://doi.org/10.1590/2176-9451.19.4.027-029.ebo. PMID: 25279518; PMCID: PMC4296634.
7. Nieminen P, Uribe SE. The quality of statistical reporting and data presentation in predatory dental journals was lower than in non-Predatory journals. **Entropy 2021 Apr 16**; 23(4): 468. https://doi.org/10.3390/e23040468. PMID: 33923391; PMCID: PMC8071575.
8. Dahiru T. P - value, a true test of statistical significance? A cautionary note. **Ann Ib Postgrad Med 2008 Jun**; 6(1): 21–26. https://doi.org/10.4314/aipm.v6i1.64038. PMID: 25161440; PMCID: PMC4111019.
9. Banerjee A, Chitnis UB, Jadhav SL, Bhawalkar JS, Chaudhury S. Hypothesis testing, type I and type II errors. **Ind Psychiatry J 2009 Jul**; 18(2): 127–131. https://doi.org/10.4103/0972-6748.62274. PMID: 21180491; PMCID: PMC2996198.
10. Haines A, Donald A. Making better use of research findings. **BMJ 1998 Jul 4**; 317(7150): 72–75. https://doi.org/10.1136/bmj.317.7150.72. PMID: 9651281; PMCID: PMC1113463.
11. Lykken DT. Statistical significance in psychological research. **Psychol Bull 1968**; 70(3, Pt.1): 151–159. https://doi.org/10.1037/h0026141.
12. Nair AS. Publication bias - Importance of studies with negative results. **Indian J Anaesth 2019 Jun**; 63(6): 505–507. https://doi.org/10.4103/ija.IJA_142_19. PMID: 31263309; PMCID: PMC6573059.

How to cite this article: AbdulRaheem Y. Statistics in medical research: Common mistakes. J Taibah Univ Med Sc 2023;18(6):1197–1199.