Journal of Statistics & Applied Sciences, Issue - 11
ISSN 2718-0999

Letter to the Editor

Received: date:20.03.2025
Accepted: date:17.06.2025
Published: date:30.06.2025

P-Hacking in Scientific Research: Is the Reliability of Scientific Results at Risk?

Sadi Elasan [1*]

[1]Department of Medical Sciences, Van Yuzuncu Yil University, Van, Türkiye; sadielasan@yyu.edu.tr
Orcid: 0000-0002-3149-6462
*Correspondence: sadielasan@yyu.edu.tr

The increasing prevalence of statistical manipulations, known as p-hacking, poses a significant challenge to the reliability of scientific findings. This issue arises when researchers repeatedly adjust data analyses until a p-value falls below 0.05, often due to publication pressure or insufficient statistical training. Such practices compromise the validity of research outcomes, particularly in fields like medicine and biomedical sciences, where false conclusions can have serious implications.

P-hacking commonly involves selective data reporting, data dredging, and multiple hypothesis testing without proper corrections, leading to inflated false positive rates [1]. For instance, when 20 independent hypothesis tests are conducted at a 5% significance level, the probability of obtaining at least one false positive result rises to approximately 64% (1-0.95^20=0.64). If 100 hypotheses are tested, this probability escalates to 99% (1-0.95^100=0.99). Small sample sizes further exacerbate this issue, with false positive rates reaching up to 50% in studies with n<30. These inflated error rates mislead researchers and clinicians, potentially affecting medical decision-making [1-2].

To mitigate p-hacking, a multifaceted approach is necessary. First, strengthening statistical training is crucial. Researchers should move beyond reliance on p-values and incorporate alternative metrics such as effect sizes, confidence intervals, and Bayesian methods. For example, a study with p<0.05 but a negligible effect size (e.g., Cohen's d=0.2) may lack real-world significance [3-5].

Second, open science practices, including preregistration and data sharing, should be widely adopted. Journals implementing open data policies, such as Psychological Science and PLOS Biology, have reported a reduction in questionable p-values, with Psychological Science observing a 40% decline in studies reporting p-values just below 0.05 (1-2-3-4).

Third, proper statistical corrections, such as the Bonferroni and Benjamini-Hochberg methods, should be routinely applied. When 20 hypotheses are tested, the Bonferroni correction adjusts the significance threshold to 0.05/20=0.0025, reducing false positive risks. Additionally, prioritizing effect size metrics like Hedges' g and reporting confidence intervals can improve result interpretation.

Finally, addressing publication bias is essential. Journals should encourage the publication of null results and studies with robust methodological designs, regardless of significance. Moreover, training programs on ethical statistical practices should be expanded, as researchers with formal statistical education are less likely to engage in p-hacking.

In conclusion, tackling p-hacking requires improved statistical training, open science initiatives, and methodological rigor. By revising editorial policies and promoting best practices, journals can play a pivotal role in preserving the integrity of scientific research.

Citation : S. Elasan, "P-hacking in scientific research: Is the reliability of scientific results at risk?" *Journal of Statistics and Applied Sciences*, no. 11, pp. 70–71, 2025. [Editor's note]

References

[1] J. P. Simmons, L. D. Nelson, and U. Simonsohn, "False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant," Psychol. Sci., vol. 22, no. 11, pp. 1359–1366, 2011, doi: 10.1177/0956797611417632.

[2] M. L. Head, L. Holman, R. Lanfear, A. T. Kahn, and M. D. Jennions, "The extent and consequences of P-hacking in science," PLoS Biol., vol. 13, no. 3, p. e1002106, 2015, doi: 10.1371/journal.pbio.1002106.

[3] D. J. Benjamin et al., "Redefine statistical significance," Nat. Hum. Behav. vol. 2, pp. 6–10, 2018, doi: 10.1038/s41562-017-0189-z.

[4] J. M. Wicherts, M. Bakker, and D. Molenaar, "Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results," PLoS ONE, vol. 6, no. 11, p. e26828, 2011, doi: 10.1371/journal.pone.0026828.

[5] B. A. Nosek, C. R. Ebersole, A. C. DeHaven, and D. T. Mellor, "The preregistration revolution," Proc. Natl. Acad. Sci. U.S.A., vol. 115, no. 11, pp. 2600–2606, 2018, doi: 10.1073/pnas.1708274114.