

References

Albers, C. (2019). The problem with unadjusted multiple and sequential

statistical testing. *Nature Communications*, 10(1), 1921.

<https://doi.org/10.1038/s41467-019-09941-0>

Bazerman, M. H., & Samuelson, W. F. (1983). I won the auction but don't

want the prize. *Journal of Conflict Resolution*, 27(4), 618–634.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S.

J., & Munafò, M. R. (2013). Power failure: Why small sample size

undermines the reliability of neuroscience. *Nature Reviews*

Neuroscience, 14(5), 365–376. <https://doi.org/10.1038/nrn3475>

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*.

Elsevier Science & Technology.

<https://ebookcentral-proquest-com.proxy.library.upenn.edu/lib/upenn-ebooks/detail.action?docID=1882849>

Danziger, M., Collazo, A., Dirnagl, U., & Toelch, U. (2022). *Balancing sensitivity*

and specificity in preclinical research (p. 2022.01.17.476585). bioRxiv.

<https://doi.org/10.1101/2022.01.17.476585>

Efron, B., & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*.

Chapman and Hall/CRC. <https://doi.org/10.1201/9780429246593>

Hesterberg, T. C. (2015). What Teachers Should Know About the Bootstrap:

Resampling in the Undergraduate Statistics Curriculum. *The American Statistician*, 69(4), 371–386.

<https://doi.org/10.1080/00031305.2015.1089789>

Hoening, J. M., & Heisey, D. M. (2001). The Abuse of Power: The Pervasive

Fallacy of Power Calculations for Data Analysis. *The American*

Statistician, 55(1), 19–24. <https://doi.org/10.1198/000313001300339897>

Johnson, P. C. D., Barry, S. J. E., Ferguson, H. M., & Müller, P. (2015). Power

analysis for generalized linear mixed models in ecology and evolution.

Methods in Ecology and Evolution, 6(2), 133–142.

<https://doi.org/10.1111/2041-210X.12306>

Krzywinski, M., & Altman, N. (2013). Power and sample size. *Nature Methods*,

10(12), 1139–1140. <https://doi.org/10.1038/nmeth.2738>

Lazic, S. E. (2022). Genuine replication and pseudoreplication. *Nature*

Reviews Methods Primers, 2(1), 23.

<https://doi.org/10.1038/s43586-022-00114-w>

Lazic, S. E., Clarke-Williams, C. J., & Munafò, M. R. (2018). What exactly is 'N' in cell culture and animal experiments? *PLOS Biology*, 16(4), e2005282.

<https://doi.org/10.1371/journal.pbio.2005282>

Reinagel, P. (2023). Is N-Hacking Ever OK? The consequences of collecting more data in pursuit of statistical significance. *PLOS Biology*, 21(11),

e3002345. <https://doi.org/10.1371/journal.pbio.3002345>

Savitz, D. A., Wise, L. A., Bond, J. C., Hatch, E. E., Ncube, C. N., Wesselink, A. K.,

Willis, M. D., Yland, J. J., & Rothman, K. J. (2024). Responding to

Reviewers and Editors About Statistical Significance Testing. *Annals of*

Internal Medicine, 177(3), 385–386. <https://doi.org/10.7326/M23-2430>

Sert, N. P. du, Ahluwalia, A., Alam, S., Avey, M. T., Baker, M., Browne, W. J.,

Clark, A., Cuthill, I. C., Dirnagl, U., Emerson, M., Garner, P., Holgate, S. T.,

Howells, D. W., Hurst, V., Karp, N. A., Lazic, S. E., Lidster, K., MacCallum, C.

J., Macleod, M., ... Würbel, H. (2020). Reporting animal research:

Explanation and elaboration for the ARRIVE guidelines 2.0. *PLOS Biology*,

18(7), e3000411. <https://doi.org/10.1371/journal.pbio.3000411>

Walters, S. J. (2004). Sample size and power estimation for studies with

health related quality of life outcomes: A comparison of four methods

using the SF-36. *Health and Quality of Life Outcomes*, 2, 26.

<https://doi.org/10.1186/1477-7525-2-26>