

# Assessing the median effect robustness in Silberzahn et al. (2018) through model specification

How analysis strategy affects analysis results

Sebastian Ploner, sploner1@sheffield.ac.uk, University of Sheffield

## Introduction

The Covid-19 pandemic has reaffirmed the need for rigours scientific research to inform decision-making on a scientific and societal level. However, particularly the social sciences have a history of inflated positive findings (Ionanidis) leading to the reproducibility crisis (Pashler & Wagenmakers, 2012). Simmons et al. (2011) demonstrated how the so-called *researcher degree of freedom* affect study results. This concept refers to the analytical choices researchers make and their induced variation in results. For instance, sample size can be highly deterministic of finding statistically significant effects. John et al. (2012) found 70% of researchers have at least once stopped their data collection based on a preliminary analysis. P-value simulations as a function of sample size showed an increased likelihood of obtaining false-positives for small sample sizes. Specifically, a sample size of ten had a 22.1% chance of being a false positive. Adding 20 observations per condition decreased the probability to 12.7% (Simmons et al., 2011). The authors did not insinuate any malicious intent per-se, but rather attributed the issue to an ambiguity of a “right way” to analyse data, and needing statistically significant results to publish. To prevent such bad practices the authors suggest requirements and guidelines for authors and reviewers, respectively.

Silberzahn et al. (2018) investigated the effects of analytical choices on results. Specifically, they had 29 teams of researchers investigate the same research question with the same dataset. Other than being part of the project there was no incentive for researchers to participate i.e., there was no ulterior motive for researcher to manipulate outcomes (e.g. wanting/needing to publish). The variation between the analytic approaches was substantial. There were 29 different analyses with 21 different combinations of covariates. Twenty teams found a statistically significant effect, and the odds-ratios ranged from 0.89 to 2.93 (median = 1.31). The authors also controlled for prior believes, experience and peer rated analysis quality, none of them accounted for the variation of results.

This study puts (psychological) research in a delicate position, because if researchers with honest intentions come to such drastically different conclusions despite starting off on the identical dataset what does this mean for the rest of studies?

Patel et al. (2015)

## **Methods**

## **Results**

## **Discussion**

## References

- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524–532. <https://doi.org/10.1177/0956797611430953>
- Pashler, H., & Wagenmakers, E.-J. (2012). Editors’ introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7(6), 528–530. <https://doi.org/10.1177/1745691612465253>
- Patel, C. J., Burford, B., & Ioannidis, J. P. A. (2015). Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *Journal of Clinical Epidemiology*, 68(9), 1046–1058. <https://doi.org/https://doi.org/10.1016/j.jclinepi.2015.05.029>
- Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahnik, S., Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M. A., Rosa, A. D., Dam, L., Evans, M. H., Cervantes, I. F., . . . Nosek, B. A. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, 1(3), 337–356. <https://doi.org/10.1177/2515245917747646>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>