

# Publication Bias in Psychology

## Third year paper proposal

Benjamin Lira

Effect size is the currency with which psychological research is interpreted. However, not all research has the same likelihood of being published. Studies with large effect sizes and statistically significant results are overrepresented in the scientific literature, leading to a distribution of effect sizes that is biased upward, a phenomenon known as publication bias. How large is this bias? I propose a meta-review of meta-analyses in psychology to estimate the distribution of true effect sizes across psychology, taking into account bias introduced by publication status.

Specifically, as a first step, I plan to analyze how meta-analysts deal with publication bias. Second, I plan to estimate the distribution of true effect sizes across psychology, taking into account bias introduced by publication status. I plan to estimate the determinants of effect sizes and publication bias, such as sample size, discipline (i.e., social vs. cognitive psychology), and the design of study (i.e., gender difference, individual difference, or experiment). Finally, I plan to use the results of the meta-review to conduct a survey of researchers to estimate how miscalibrated they are in judging effect sizes and publication bias. Such contribution would allow psychologists to better calibrate their expectations of effect size when planning and reviewing research.

What do we know about effect sizes and publication bias in psychology? First, original published effect sizes are biased upward, and are roughly twice as large as careful replications. The many labs project<sup>1</sup> showed that the average effect size across 100 published experiments was  $r = .40$ , whereas the average effect size across 36 replications was  $r = .20$ . Similarly, evidence from eleven years of student replication projects, also shows that published effect sizes ( $r = .30$ ) are roughly twice the size of replications ( $r = .34$ )<sup>2</sup>. Consistent with this, unpublished effects are, on average .18 SDs smaller than their published counterparts<sup>3</sup>.

Second, efforts at characterizing median effect sizes have been circumscribed to particular sub-fields. For example, the median effect size has been estimated at  $r = .42$  for cognitive neuroscience and psychology<sup>4</sup>, at  $r = .19$  for individual differences<sup>5</sup>, and at  $r = .16$  for applied psychology<sup>6</sup>.

Third, evidence suggests that small effect sizes are to be expected even with plenty of data and sophisticated methods. Most notably, Salganik et al<sup>7</sup> conducted a prediction contest

where participants were asked to predict the life outcomes of 1000 people based on a large set of predictors. The best model was only able to explain 20% of the variance in two of the outcomes, and 5% on the other four outcomes. Moreover, The best models barely outperformed a simple linear regression model with only four predictors, suggesting that more data or more predictors is not guaranteed to improve prediction.

Finally, there is a growing consensus that small effect sizes in psychology should be expected and can still be relevant. Ahadi<sup>8</sup> conducted a simulation study that shows that when outcomes are multiply determined (an assumption that likely holds for behavior), the correlation of any one trait predictor is bounded (e.g., at .50 if the behavior is influenced by three traits). Similarly, Funder and Ozer<sup>9</sup> argued that small effects should be interpreted contextually: if a small effect affects a large numbers of people for a long time, the cumulative change could be greater than a more circumscribed but larger effect. Finally, Götz<sup>10</sup> argues that if small effect sizes are the norm rather than the exception, psychology can only be a cumulative science if we adjust our expectations regarding effect size.

## Proposed research

### Study 1

We survey the meta-analysis literature in psychology in the past 5 years. From that search, we find over 5K published meta-analyses, of which 1.2K have accessible pdfs. We can computationally analyze the text in these files to learn how meta-analysts deal with publication bias. Preliminary findings, suggest that less than half of meta-analyses conduct an inferential test for publication bias, and fewer than a quarter adjust their findings to account for publication bias.

### Study 2a

We survey pre-registered meta-analyses with existing data posted online. When publication status is reported, we compute the empirical distributions of effect size conditional on publication status. We estimate the file drawer effect size from this. Across all datasets, we apply Robust Bayesian Meta-Analysis (RoBMA)<sup>11</sup> to estimate the distribution of true effect sizes. This allows us to compare the distribution of corrected effect size to that of published effect sizes. We estimate how moderators of effect size relate to the file drawer effect size and the correction factor.

### Study 2b

A replication of Study 2, but with a different sample of meta-analyses. We test whether the results of study 2a replicate when estimated from a sample of meta-analyses provided by authors, as opposed to being publicly available.

## Study 3

Given the results of Study 2, we survey researchers who have not conducted meta-analyses to estimate the average correction factor.

## Research questions

- How do meta-analysts analyze publication bias in the past 5 years of published meta-analyses?
- What is the distribution of true effect sizes across psychology?
- How do moderators of effect size relate to the file drawer effect size and the correction factor?
- How miscalibrated are researchers in judging effect sizes and publication bias?

## Methods

### Study 1

We plan to download all meta-analyses published in the past 5 years in psychological journals. We will then download all available pdfs, and analyze them using natural language processing tools (i.e., dictionary methods, text classifiers) to extract the following information:

1. What percentage of meta-analyses collect unpublished effect sizes?
2. What percentage of meta-analyses conduct an inferential test for publication bias?
3. What percentage of meta-analyses adjust their findings to account for publication bias?
4. How do these trends change over time.

### Study 2

We plan to obtain all meta-analyses with existing data posted online. From these we can estimate two distribution of effect sizes: one correcting for publication bias, and one uncorrected. From this, I will estimate the average correction factor on published effect sizes. Finally, I plan to use information on the characteristics of the meta-analyses to estimate the determinants of effect sizes and publication bias (i.e., moderation analyses).

### Study 3

We will conduct a survey of researchers to estimate how miscalibrated they are to true effect sizes. We will use the results of the meta-review to create a set of vignettes that will be used to estimate miscalibration.

## Reading List

1. Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
2. Boyce, V., Mathur, M. B., & Frank, M. C. (2023). Eleven years of student replication projects provide evidence on the correlates of replicability in psychology [Preprint]. *PsyArXiv*. <https://doi.org/10.31234/osf.io/dpyn6>
3. Polanin, J. R., Tanner-Smith, E. E., & Hennessy, E. A. (2016). Estimating the Difference Between Published and Unpublished Effect Sizes: A Meta-Review. *Review of Educational Research*, 86(1), 207–236. <https://doi.org/10.3102/0034654315582067>
4. Szucs, D., & Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLOS Biology*, 15(3), e2000797. <https://doi.org/10.1371/journal.pbio.2000797>
5. Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, 102, 74–78. <https://doi.org/10.1016/j.paid.2016.06.069>
6. Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. *Journal of Applied Psychology*, 100(2), 431–449. <https://doi.org/10.1037/a0038047>
7. Salganik, M. J., Lundberg, I., Kindel, A. T., Ahearn, C. E., Al-Ghoneim, K., Almaatouq, A., Altschul, D. M., Brand, J. E., Carnegie, N. B., Compton, R. J., Datta, D., Davidson, T., Filippova, A., Gilroy, C., Goode, B. J., Jahani, E., Kashyap, R., Kirchner, A., McKay, S., ... McLanahan, S. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences*, 117(15), 8398–8403. <https://doi.org/10.1073/pnas.1915006117>
8. Ahadi, S., & Diener, E. (1989). Multiple determinants and effect size. *Journal of Personality and Social Psychology*, 56(3), 398–406. <https://doi.org/10.1037/0022-3514.56.3.398>
9. Funder, D. C., & Ozer, D. J. (2019). Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science*, 2(2), 156–168. <https://doi.org/10.1177/2515245919847202>
10. Götz, F. M., Gosling, S. D., & Rentfrow, P. J. (2022). Small Effects: The Indispensable Foundation for a Cumulative Psychological Science. *Perspectives on Psychological Science*, 17(1), 205–215.
11. Maier, M. (2023). Robust Bayesian Meta-Analysis: Addressing Publication Bias With Model-Averaging.