

Small effects are a big deal

Third year paper proposal

Benjamin Lira

Effect size is the currency with which psychological research is interpreted. Despite its importance, there is little consensus on what effect size a researcher can expect when conducting a study. Because of publication bias, the effect sizes that are reported in the literature are biased upward, relative to what a researcher should expect a priori. Moreover, traditional effect size benchmarks (e.g., Cohen's $r = .10$, $.30$, and $.50$ for small, medium, and large correlations, respectively,¹) do not take into account the fact that effect sizes can vary widely depending on the type of study. For example, a longitudinal study where an intervention is supposed to change a real life outcome might have far smaller effects than a cross sectional study where self-reported attitudes are correlated with each other. Fadeout², common method variance, and study design are just some of the factors that can influence effect size.

To address these limitations, I propose a meta-review of meta-analyses in psychology. Specifically, I plan to estimate the distribution of true effect sizes across psychology, taking into account bias introduced by publication status. Second, I plan to estimate the determinants of effect sizes, such as study design, sample size, and publication status. Additionally, 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 whether this miscalibration is specific. That is, do researchers uniformly overestimate effect sizes, or do they fail to account for the determinants of effect size? 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 in psychology? First, there is a growing consensus that small effect sizes in psychology should be expected and can still be relevant. Ahadi³ conducted a simulation study to show 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 Ozer⁴ 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, Götz⁵ 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.

Second, original published effect sizes are biased upward, and are roughly twice as large as careful replications. The many labs project⁶ 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 ($r = .34$)⁷. Consistent with this, unpublished effects are, on average .18 SDs smaller than their published counterparts⁸.

Third, even with plenty of data and sophisticated methods, predicting life outcomes is remarkably difficult. Most notably, Salganik⁹ conducted a prediction contest where participants were asked to predict the outcomes of 1000 people based on a large set of predictors. The best model was only able to explain 20% of the variance in the two of the outcomes, and 5% on the other four outcomes. Notably, 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, 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 psychology¹⁰, at $r = .19$ for individual differences¹¹, and at $r = .16$ for applied psychology¹².

Research questions

- What is the distribution of published and unpublished effect sizes in psychology?
- How do the following features of studies relate to effect size?
 - Lab studies vs. field studies
 - Experimental vs. correlational studies
 - Social vs. cognitive
 - Within person vs. between person
 - Concurrent vs. time-lagged
 - Common method vs multi-method
- How miscalibrated are scientists to true effect sizes. For what kinds of research are they more likely to overestimate effect sizes?

Methods

Meta-review

We plan to download all meta-analyses published in the past 5 years in psychological journals. Because not all meta-analyses report published and unpublished effect sizes, we will email researchers request their meta-analytic datasets. I will also ask authors to specify

the predictor, criterion, and field of their study. Then, I will aggregate these meta-analytic datasets, and use meta-regression to estimate the distribution of published and unpublished effect sizes, and to test for the moderators of effect size.

Survey

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.

Progress so far

I have downloaded 5,474 meta-analyses from the past 5 years. Of these, 3,567 have accessible pdfs, from which I am extracting emails. I have created a [survey](#) to ask authors to upload their data and categorize their meta-analysis.

Reading List

1. Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
2. Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and Fade-Out of Educational-Intervention Effects: Mechanisms and Potential Solutions. *Psychological Science in the Public Interest*, 21(2), 55–97. <https://doi.org/10.1177/1529100620915848>
3. 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>
4. 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>
5. Götz, F. M., Gosling, S. D., & Rentfrow, P. J. (2022). Small Effects: The Indispensable Foundation for a Cumulative Psychological Science.
6. Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
7. 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>
8. 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>
9. 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>
10. 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>
11. 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>

12. 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>