

Small effects are a big deal

Third year paper proposal

Benjamin Lira

There is a lot that has been said about effect sizes in psychology. Since Cohen¹ published effect size benchmarks of $r = .10$, $.30$, and $.50$ for small, medium, and large correlations, respectively, recent inspection of true effect sizes in psychology has revealed that these original guidelines might have been unrealistic. There are four converging sources of evidence that point to this conclusion.

The first such avenue has relied on theoretical reasons, based on reasoning or simulation, to argue that small effects matter, and that small effects are to be expected. For example, Ahadi² conducted a simulation study to show that for outcomes that 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 Ozer³ 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, Götz⁴ 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, large scale replication projects have shown that published effect sizes are biased upward. The many labs project⁵ showed that the average effect size across 100 published experiments was $r = .40$, whereas the average effect size across 36 replications was $r = .20$. Interestingly, cognitive psychology had higher effect sizes than social psychology.

Third, prediction contests suggest that even with plenty of data and sophisticated methods, predicting life outcomes is remarkably difficult. Most notably, Salganik⁶ 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, evidence from meta-analyses, especially for particular sub-fields, provide some indication of what effect sizes might be expected (see for example summary effect sizes in cognitive neuroscience⁷, nudges⁸, individual differences⁹, or applied psychology¹⁰).

Despite this, there is still ample miscalibration among researchers on what are credible and practically significant effect sizes. What's more, most effect sizes are biased upward due to publication bias. Polanin¹¹ conducted a meta-review to estimate the file drawer effect, and found that unpublished effects are, on average .18 SDs smaller than their published counterparts. Contrary to this, Dalton et al¹² compared unpublished dissertations with journal articles to argue that there is no file drawer inflation bias in personnel psychology.

Not only are researchers miscalibrated in what effect sizes are to be expected, but also, there is no evidence available to help researchers understand where effect sizes come from. There is reason to believe that longitudinal effects are probably smaller than concurrent ones, because of fadeout¹³. Common method variance might also inflate effect sizes. The ecological validity of a study might also affect effect sizes.

We propose a meta-review of effect sizes across psychology to address these limitations. Specifically, we want to estimate the distribution of true effect sizes across psychology, taking into account bias introduced by publication status. We expect great heterogeneity in effect sizes, and thus also plan to explain this heterogeneity: what are the determinants of effect size. Such contribution would allow psychologists to better calibrate their expectations of

effect size when planning and reviewing research.

Additionally, we 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?

Research questions

- What is the distribution of effect sizes in psychology?
- How does this distribution change for different types of studies?
 - Published vs. unpublished
 - 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

Our plan is to conduct a meta-review of meta-analyses in psychology. We will use meta-analytic effect sizes as our dependent variable. We plan to obtain a representative sample of meta-analyses in psychology from the past 10 years. We will code for the moderators above,

as well as the predictor, the criterion, the average sample size per study, and the average number of studies per meta-analysis, and the average age of participants.

We will use meta-regression to estimate the distribution of 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.

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. 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>
3. 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>
4. Götz, F. M., Gosling, S. D., & Rentfrow, P. J. (2022). Small Effects: The Indispensable Foundation for a Cumulative Psychological Science.
5. Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
6. 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>
7. 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>
8. DellaVigna, S., & Linos, E. (2022). RCTs to Scale: Comprehensive Evidence From Two Nudge Units. *Econometrica*, 90(1), 81–116. <https://doi.org/10.3982/ECTA18709>
9. 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>
10. 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>
11. 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>
12. Dalton, D. R., Aguinis, H., Dalton, C. M., Bosco, F. A., & Pierce, C. A. (2012). Revisiting the file drawer problem in meta-analysis: An assessment of published and nonpublished correlation matrices: Personnel

- Psychology. *Personnel Psychology*, 65(2), 221–249. <https://doi.org/10.1111/j.1744-6570.2012.01243.x>
13. 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>

SCRAPS

Open questions

- Is this a tractable problem?
- If it were possible, would it be a useful contribution?
- We expect to have a missing data problem with some of the moderators of effect size. How big of a problem would this be?
- Given what already is out there, is it true that this is a case of “go big or go home”?
- Would it be useful to also use metaBUS for this?

Notes from Meetings

- Dolores thinks that just using Psych Bulletin data is not gonna fly in Psych Bulletin. It could work for PNAS or some other journal she said.
- An interesting point from Dolores was to use the estimated distribution to try to uncover insights from psychology. To interesting examples: (1) Are effects larger for younger samples? Are kids more malleable? (2) Are effects larger the closer you get to behavior.