Boundary conditions for the practical importance of small effects in long-runs: A comment on Funder and Ozer (2019).

James D. Sauer [a,b*], and Aaron Drummond [b,c]

[a] School of Psychological Sciences, University of Tasmania, Australia

[b] International Media Psychology Laboratory

[c] School of Psychology, Massey University, Palmerston North 4424, Manawatu, New Zealand

[*]To whom correspondence should be addressed. Email: Jim.Sauer@utas.edu.au

Funder and Ozer (2019) argued that small effects can have important implications in cumulative long-run scenarios. We certainly agree. However, some important caveats merit explicit consideration. We elaborate on the previously-acknowledged importance of pre-registration (and open data practices), and identify two additional considerations for interpreting small effects in long-run scenarios: restricted extrapolation and construct validity.

Pre-Registration and Open Data

Interpreting small effects requires discrimination at two levels. First, discriminating reliable-but-small effects from null effects where flexibility in analysis has inflated effect size. Second, distinguishing between reliable-but-small effects that are and not likely to cumulate to meaningfully predict or affect behaviour. Analysis involves a series of decisions – covariate inclusion/exclusion, sub-scale analysis, outlier treatment, etc. – known collectively as *the garden of forking paths* (Gelman & Loken, 2014). Researcher degrees of freedom can increase Type-1 error rates and inflate effect sizes (Simmons, Nelson, & Simonsohn, 2011). Pre-registering analysis plans constrains these forking paths. Flexibility in analysis can inflate effects of all sizes, but the consequences of such inflation – and associated risks of over-interpreting effects and over-estimating the likelihood they will cumulate to produce practically-important outcomes – are perhaps greater if inflation nudges an effect size from below to above the cut-off for the “smallest effect size of interest” (rather than, for example, from moderate to large). Studies reporting small effects but lacking or deviating from pre-registered analysis plans should be interpreted cautiously. Further, although not a panacea for research bias, open data practices help establish the robustness of small effects in two ways. First, effects contingent on highly-contrived analysis strategies should be interpreted and generalized with caution. Second, open data practices allow interested researchers to estimate the “smallest effect of interest” by, for example, comparing the small effects of interest with nonsense relationships in the data (Orben & Pryzbylski, 2019).

Restricted Extrapolation

To illustrate the importance of small effects, Funder and Ozer (2019) present examples where small effects accumulate to produce meaningful outcomes. First, they consider the small correlation between a baseball player's performance for any individual at-bat and batting average (Abelson, 1985). This context demonstrates the importance of small effects: single trials cumulatively contribute to players' batting averages and, subsequently, teams' winning percentages. It also highlights some important caveats. First, the relationship between single at-bats and batting average is the relationship between A and Ā across a series of trials. Such measurement purity is uncommon in psychology and increasing measurement noise decreases the predictive value of small effects. Second, we must be cautious extrapolating small observed relationships to third variables. Although at-bat performance cumulatively relates to batting averages (because it *indexes* cumulative performance over single at-bats), this small relationship is, relatively speaking, not a particularly useful predictor of *team* performance. A player is valuable *because he gets on base*: On-base percentage contributes twice as much as batting averages to winning percentage (Hakes & Sauer, 2006). Small effects may cumulate to produce reliable relationships, but the predictive value of these relationships for third variable outcomes in practical settings may remain low and, therefore, *relatively* unimportant.

Similarly, a small relationship between ego depletion and having a "short-fuse" in stressful conversations may predict a slightly higher likelihood of disagreements *during* stressful conversations, without necessarily translating to third-variable outcomes (e.g., marital friction; Funder & Ozer, 2019).

Even in causal relationships, small effects of A on B do not guarantee meaningful third-variable consequences. Relationships between second and third variables in causal chains are usually imperfect, plausibly constraining indirect effects to potentially inconsequential levels. We must be cautious not to overstate the importance of small effects by extrapolating to unmeasured consequences. Small effects of A on B may cumulatively produce meaningful changes in B, without implying meaningful effects on C. Further, as Funder and Ozer noted, current theorizing typically precludes robust predictions about whether small effects will cumulate in strength or consequences.

Construct Validity

The example of batting averages and performance at-bat is almost unique. Batting performance is not operationalized, it is a *direct measure* of the target construct. Thus, the example translates poorly to most psychological research where interests often relate to latent constructs (e.g., personality, memory). Even examining observable behaviours (e.g., cigarettes smoked) often requires indirect measurement (e.g., self-report). Relationships between constructs and operationalized variables are imperfect, and when constructs are unobservable, the magnitude and direction of measurement error is unknowable. Thus, Funder and Ozer's pre-condition that estimation is reliable may often be unverifiable. A small effect on an operationalized variable (r = .05) may be stronger or weaker than the effect on the construct of interest (Cohen, 1988, 1992). For example, manipulating a salient variable in tightly controlled experimental conditions may exaggerate effects relative to complex applied settings (Schäfer, & Schwarz, 2019), demand artifacts may inflate effects (Sawyer, 1975), or effects on self-reported outcomes may overstate effects on *actual* outcomes. When true effects are weaker than small observed effects, they may be statistically equivalent to zero. Thus, small effects should be interpreted with greater caution as measurements become less direct and constructs more abstracted.

Moreover, the relationship between at-bat performance and batting averages contains few plausible confounds. This is untrue for most psychological constructs and, in complex systems, small effects may easily be produced or mitigated by uncontrolled confounds. For instance, regarding ego depletion, perhaps someone hates their job because they are an anxious or aggressive person, and these traits (not self-control depletion), explain their "shorter fuse". Alternatively, good communication may practically eliminate the small effect of depleted ego on disagreements.

Concluding Remarks

Small effects *may* have important cumulative impacts over long-runs. However, the boundary conditions for such cumulative effects are poorly understood. Pre-registration and open data will likely improve verifiability of small effects, but our ability to reliably determine the practical importance (i.e., cumulative consequences) of small effects remains under-developed. Where plausible uncontrolled confounds exist, indirect measurement is employed, or constructs of interest are latent, researchers should be cautious when interpreting the practical importance of small effects. Similarly, researchers should limit interpretations to available data, not extrapolate small effects to unobserved third variables.

Author contributions

Both authors contributed equally to all aspects of this work.

Conflicts of Interest

The authors declare that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

Supported by the Marsden Fund Council from Government funding, managed by Royal Society Te Apārangi; MAU1804.

References

Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin, 97*(1), 129.

Cohen, J. (1988). Statistical power analysis for the behavioural sciences: Hillsdale, NJ: erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155.

Gelman, A. , & Loken, E. (2014). The statistical crisis in science. *American Scientist, 102*, 460.

Hakes, J. K., & Sauer, R. D. (2006). An economic evaluation of the Moneyball hypothesis. *Journal of Economic Perspectives, 20*(3), 173-186.

Orben, A., & Przybylski, A. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behavior, 3*, 173-182.

Sawyer, A. G. (1975). Demand artifacts in laboratory experiments in consumer research. *Journal of Consumer Research*, *1*, 20-30.

Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology*, *10*, 813.

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.

St. John, A. (2013). Powered by Bill James and friends, the Red Sox win (another) Moneyball World Series. *Forbes*. Retrieved from https://www.forbes.com/sites/allenstjohn/2013/10/31/powered-by-bill-james-and-friends-the-red-sox-win-another-moneyball-world-series/#4f2362947c64