Supplementary Material

Statistical methods for comparing meta-analytic and MLR estimates

In the Main Text, we report the correlation between the meta-analytic and MLR effect sizes estimates (Pearson's $r(13) = 0.72$ [0.32, 0.90], $p = 0.003$). This estimate reflects the correlation between the random effect meta-analytic estimates and the MLR effect size estimates. The correlation is similar when the original meta-analytic effect sizes estimates are used ($r(13) = 0.72$ [0.34, 0.9], $p = 0.002$). It is also similar when you exclude the one meta-analysis for which an estimate of effect size in terms of Cohen's d is not available (Srull & Wyer, 1979; $r(12) = 0.72$ [0.3, 0.9], $p = 0.004$). Code for all analyses can be found in the associated repository: https://github.com/mllewis/kvarven_reanalysis.

Statistical methods for sensitivity analyses

Technical details on the worst-case sensitivity analysis method are provided elsewhere (1); here we provide a brief intuitive overview. The method considers a mechanism of publication bias in which "statistically significant" and positive studies (henceforth "affirmative studies") are more likely to be published and included in the meta-analysis than "nonsignificant" or negative studies (henceforth "nonaffirmative studies") and that does not select further based on the size of the point estimate or on characteristics associated with the point estimate (such as the p-value treated as continuous, rather than dichotomized at $\alpha = 0.05$). This model of publication bias is identical to that assumed by the three-parameter selection model (2) used by Kvarven et al. (3), and it conforms well to empirical evidence regarding how publication bias operates in practice (1).

Under this model of publication bias, a bias-corrected meta-analytic estimate could hypothetically be obtained by upweighting the contribution of each nonaffirmative study in the meta-analysis by the same ratio by which the publication process favors *affirmative* studies, if this ratio were known. For example, suppose it were known that affirmative results were five times more likely to be published than nonaffirmative studies and that, given a study's nonaffirmative or affirmative status, the publication process did not select further based on the size of the point estimate. Then the point estimates of the published nonaffirmative studies (i.e., those included in the meta-analysis) would be essentially a random sample of those from the larger, underlying population of nonaffirmative studies, most of which are not published. A bias-corrected meta-analytic estimate could therefore be obtained by upweighting the contribution of each nonaffirmative study in the meta-analysis by fivefold to counteract the publication process' fivefold favoring of affirmative studies. By extension, worst-case publication bias would effectively favor affirmative studies by an infinite ratio, so a worst-case estimate can be obtained by meta-analyzing only the nonaffirmative studies that are included in the meta-analyses and simply discarding the affirmative studies; this works because such an analysis is effectively equivalent to upweighting each nonaffirmative study by a factor of infinity (1). Such a worst-case analysis does not require actual estimation of the ratio by which the publication process favors affirmative studies.

Standard selection models (e.g., 2) attempt to statistically estimate this ratio, but in doing so, they require assumptions regarding the distribution of true population effects prior to selection by publication bias and typically work well only in relatively large meta-analyses (e.g., at least 30-40 studies). Because the sensitivity analyses retain the same model of publication bias as these selection models, but do not require these other assumptions, they can help clarify whether the apparent failure of selection models to adequately bias-correct meta-analyses when MLR estimates are taken as the gold standard (3) reflects a genuine failure of the selection models due to violated assumptions or rather indicate that the MLR estimates are not, in fact, the appropriate gold standard against which to compare meta-analyses.

(1) Mathur, M. B., & VanderWeele, T. J. (under review). Sensitivity analysis for publication bias in meta-analyses. Preprint retrieved from: https://osf.io/s9dp6/.

(2) Vevea, J. L., & Hedges, L.. A general linear model for estimating effect size in the presence of publication bias." *Psychometrika* 60.3 (1995): 419-435.

(3) Kvarven, A., Strømland, E., & Johannesson, M. Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour*, 1-12 (2019)