**Using Selection Models to Assess Sensitivity to Publication Bias:
A Tutorial and Call for More Routine Use**

Maximilian Maier^{1,2*}, Tyler J. VanderWeele³, and Maya B. Mathur⁴

1 Department of Experimental Psychology, University College London

2 Department of Psychology, University of Amsterdam

3 Department of Epidemiology, Harvard University

4 Quantitative Sciences Unit and Department of Pediatrics, Stanford University

*Correspondence concerning this article should be addressed to:

Maximilian Maier

University College London

Department of Experimental Psychology

26 Bedford Way 129-B

WC1H 0AP London

United Kingdom

maximilian.maier.20@ucl.ac.uk

This preprint has not been peer reviewed. It is therefore not the authoritative document of
record.

Abstract

In meta-analyses, it is critical to assess the extent to which publication bias might have compromised the results. Classical methods based on the funnel plot, including Egger’s test and Trim-and-Fill, have become the de facto default methods to do so, with a large majority of recent meta-analyses in top medical journals (85%) assessing for publication bias exclusively using these methods. However, these classical funnel plot methods have important limitations when used as the sole means of assessing publication bias: they essentially assume that the publication process favors large point estimates for small studies and does not affect the largest studies, and they can perform poorly when effects are heterogeneous. In light of these limitations, we recommend that meta-analyses routinely apply other publication bias methods in addition to or instead of classical funnel plot methods. To this end, we describe how to use and interpret selection models. These methods make the often more realistic assumption that publication bias favors “statistically significant” results and that also directly accommodate effect heterogeneity. Selection models are well-established in the statistics literature and are supported by user-friendly software, yet remain rarely reported in many disciplines. We use previously published meta-analyses to demonstrate that selection models can yield insights that extend beyond those provided by funnel plot methods, suggesting the importance of establishing more comprehensive reporting practices for publication bias assessment.

Keywords: Heterogeneity, Funnel Plot, Egger Regression, Meta-Analysis

Using Selection Models to Assess Sensitivity to Publication Bias: A Tutorial and Call for More Routine Use

1 In meta-analyses, publication bias – such as the preferential publication of papers
2 supporting a given hypothesis rather than null or negative results – can lead to incorrect
3 meta-analytic estimates of the mean effect size. Among the most popular tools to assess
4 and correct for publication bias is the funnel plot (e.g., Duval & Tweedie, 2000; Egger
5 et al., 1997), a graph relating the meta-analyzed studies’ point estimates to a measure of
6 their precision, such as sample size or standard error. Funnel plots are usually interpreted
7 as indicating publication bias when they are asymmetric, that is, when smaller studies tend
8 to have larger point estimates. Several popular statistical methods, such as Trim-and-Fill
9 (e.g., Duval & Tweedie, 2000) and Egger’s regression (Egger et al., 1997) are designed to
10 quantify this type of asymmetry. When referring to “funnel plot methods”, we focus on
11 these classical, very widespread approaches rather than on recent extensions that have not
12 yet become common practice (Bartoš et al., 2021; Stanley & Doucouliagos, 2017; Stanley
13 et al., 2017).

14 Among meta-analyses in medicine that included some assessment of publication
15 bias, nearly all do so using classical funnel plot methods. We reviewed meta-analyses
16 published in *Annals of Internal Medicine*, *Journal of the American Medical Association*,
17 and *Lancet*. To do so, we systematically sampled 25 meta-analyses by journal (75 total),
18 reviewing the meta-analyses reverse-chronologically from 2019 back to 2017. (More details
19 about the methodology and results of the review are available in the Online Supplement.)
20 Among these meta-analyses, 55% did not assess publication bias at all. Of the 45% that
21 did assess publication bias, 85% did so using classical methods based on the funnel plot,
22 such as visual inspection of funnel plots, Egger’s regression, or Trim-and-Fill. The
23 remaining 15% used methods such as the fail-safe N (Rosenthal, 1986) or the test for excess
24 significance (Ioannidis & Trikalinos, 2007). Unfortunately, none of these 75 meta-analyses

used selection models, an important and methodologically viable alternative that we discuss below.

The Need for Other Methods: Uses and Limitations of Funnel Plot Methods

Funnel plot methods can be useful to assess general “small-study effects” (Sterne et al., 2011): that is, whether the effect sizes differ systematically between small and large studies. Such effects could arise not only from publication bias but also from genuine substantive differences between small and large studies (Egger et al., 1997; Lau et al., 2006). For example, in a meta-analysis of intervention studies, if the most effective interventions are also the most expensive or difficult to implement, these highly effective interventions might be used primarily in the smallest studies. Funnel plot methods detect these types of small-study effects as well as those arising from publication bias.

In practice, though, funnel plot methods are often used and interpreted specifically as means of assessing publication bias rather than as means of assessing these general small-study effects. Thus, in this context of publication bias assessment, these methods have important limitations. First, they effectively assume that small studies with large positive point estimates are more likely to be published than small studies with small or negative point estimates. Second, typically, they effectively assume that the largest studies are published regardless of the magnitude of their point estimates (Rothstein et al., 2005, pp.75-99). This is because if publication bias does indeed operate in this manner, then in a meta-analysis without publication bias, larger studies would cluster more closely around the true mean (i.e., the mean of all studies, whether published or unpublished) than smaller studies, but large and small studies alike would have point estimates centered around the true mean (Borenstein et al., 2011, p.283). Thus, the point estimates would tend to form a symmetric “funnel” shape. In a meta-analysis more severely affected by publication bias, the reasoning goes, small studies with small or negative point estimates would more frequently be omitted from the plot than small studies with large positive point estimates

or large studies. This selective publication would lead to an asymmetric funnel shape in which the observed small studies tend to have larger point estimates than larger studies.

As our review indicates, funnel plot methods essentially remain the sole means of assessing publication bias in a large majority of high-profile medical meta-analyses, but echoing others' caveats (Lau et al., 2006; Sterne et al., 2011), we believe that this exclusive focus is problematic. As noted above, when funnel plot methods are interpreted as indications of and corrections for publication bias rather than as indications of general small-study effects, the methods make implicit assumptions about how publication bias operates and about the distribution of effect sizes across studies. These assumptions are rarely stated in papers that apply funnel plot methods, and in many meta-analyses, the assumptions may not align well with the way publication bias operates in practice.

Specifically, funnel plot methods assume that publication bias operates on the size of studies' point estimates regardless of their p -values, yet empirical evidence suggests that selection for "statistically significant" p -values less than 0.05 is in fact a strong influence and likely stronger than selection for point estimate size (Masicampo & Lalande, 2012; McShane & Gal, 2017; Nelson et al., 1986; Rosenthal & Gaito, 1963, 1964; Wicherts, 2017). For better or for worse, in most fields, it is "significant" p -values that researchers attempt to achieve and that attract the attention of peer reviewers and editors. Furthermore, in many meta-analyses, it seems implausible that large studies are immune to publication pressures, a situation that further violates the assumptions of funnel plot methods. Funnel plot methods may fail to detect publication bias that operates on p -values if such publication bias does not induce a strong correlation between studies' estimates and standard errors, or if publication bias selects for "significant" results in either direction.^a In either of these cases, publication bias might not induce asymmetry in the funnel plot.

^a Publication bias that favors "significant" results with positive estimates. does induce some degree of association between estimates and standard errors, but this association is nonlinear (Stanley & Doucouliagos, 2014). Because funnel plot methods specifically detect *linear* correlations whereas selection models more precisely capture the nonlinear form of the association that arises under this type of

Besides their assumptions regarding the publication process, funnel plot methods can perform poorly when effects are heterogeneous across studies, as we detail below. For these reasons, funnel plot methods should be supplemented with selection models, or other comparable methods, so as to more effectively assess publication bias.

Selection Models as an Additional Means to Assess Publication Bias

We are not opposed to using funnel plot methods for assessing general small-study effects, but it is problematic to use them as the sole means of assessing publication bias, as is current practice in high-profile medical meta-analyses. In light of the limitations of funnel plot methods, we believe that meta-analyses should routinely apply other methods as well. As one reasonable alternative, selection models are statistically well-established methods (Hedges, 1984; Iyengar & Greenhouse, 1988; Vevea & Hedges, 1995) and make more flexible assumptions regarding publication bias and heterogeneity that we believe are more realistic for the majority of meta-analyses. For example, selection models can be specified to allow for publication bias that favors studies with “statistically significant” p -values less than 0.05 and with point estimates in the desired direction, regardless of their sample sizes or the magnitude of their point estimates (Veeva & Hedges, 1995). Selection models of this form essentially assess whether, among the published studies, there is a relative overrepresentation of studies with significant p -values and positive estimates compared to studies with nonsignificant p -values or negative estimates. To do so, the models use maximum likelihood estimation to obtain a bias-adjusted meta-analytic mean by giving more weight to observed studies that are underrepresented in the sample due to publication bias (i.e., those with point estimates in the unexpected direction or with “nonsignificant” p -values; Card, 2015, pp. 274-275). Additionally, selection models directly

publication bias that selects for statistical “significance”, selection models often have greater statistical power to detect publication bias (Pustejovsky & Rodgers, 2019).

accommodate effect heterogeneity through the specification of, for example, a normal likelihood for the effect distribution (Rothstein et al., 2005; Terrin et al., 2003).

Example - Funnel Plot Methods and Selection Models Can Provide Different Insights

The differing assumptions of funnel plots versus selection models are not merely a point of statistical pedantry but rather can lead the methods to provide differing insights when applied to published meta-analyses. For example, Toosi et al. (2012) meta-analyzed the effect of interracial interactions on positive attitudes, negative affect, nonverbal behavior and performance. A standard random effects model, prior to any correction for publication bias, indicates that performance was somewhat higher in same race pairs compared to dyads of different races, $r = 0.07$ (95% confidence interval [CI]: [0.02, 0.12], $p = 0.003$) with estimated heterogeneity $\hat{\tau} = 0.15$ (95% CI: [0.12, 0.21]).^b They assessed publication bias using Egger’s regression, finding little evidence for funnel plot asymmetry. In addition, visual inspection of the funnel plot does not seem to indicate publication bias (Figure 2).

We reanalyze the data with a simple selection model that assumes that “significant” studies, regardless of effect direction, are more likely to be published than “nonsignificant” studies. Our re-analysis suggests evidence for publication bias, $\chi^2(2) = 11.82$, $p = 0.003$. Accordingly, the bias-corrected meta-analytic mean is smaller and has a larger p -value, $r = 0.04$ (95% CI: [-0.02, 0.09], $p = 0.087$). As noted above, these differing results could occur if, for example, there is little correlation between studies’ estimates and standard errors (as detected by Egger’s regression), yet there is an apparent overrepresentation of “significant” studies in either direction (as detected by the selection model). This example illustrates that using selection models to accommodate a more realistic mechanism of publication bias

^b For all analyses, we converted correlations to Fisher’s Z scale and transformed them back for reporting; however, the heterogeneity is reported on Fisher’s Z scale directly.

and effect heterogeneity can result in different insights from those provided by funnel plot methods.

Figure 1

Funnel Plot for Toosi et al. (2010)

Note. Correlations (Observed Outcome) are displayed after Fisher’s Z transformation.

Practical Recommendations for Applying and Interpreting Selection Models

Selection models are easy to implement in practice. The R package `weightr` includes a function to fit a selection model as follows (Coburn et al., 2019):

```
library(weightr)

weightfunct(effect, v)
```

(1)

where “effect” represents the studies’ point estimates and “v” represents their estimated variances (i.e., squared standard errors). By default, the `weightr` package assumes that studies with a two-tailed $p < 0.05$ and with positive estimates have a higher

probability of publication than studies with $p \geq 0.05$ or with negative estimates. Therefore, if the hypothesized direction of the effect is negative rather than positive, the signs of all point estimates should first be reversed before using this code. In this model of publication bias, effects in a particular direction are favored for publication. However, it is possible in some contexts that all “significant” results are favored regardless of direction; to fit a more flexible selection model that accommodates this possibility, the R code in code excerpt (1) could be modified as follows:

```
weightfunct(effect, v, steps = c(0.025, 0.975, 1))
```

(2)

The **steps** argument represents the p -value cutoff points, representing the p -values at which a result’s probability of being published is thought to change. To facilitate specifying publication bias that selects for the estimate direction as well as the p -value, the **steps** argument is specified in terms of one-tailed, not two-tailed, p -values. We refer to one-tailed p -values as “ p_+ ” to distinguish them from two-tailed p -values. That is, a one-tailed $p_+ < 0.025$ represents a positive estimate that is “significant” on a *two*-tailed test, whereas a one-tailed $p_+ > 0.975$ represents a negative estimate that is “significant” on a two-tailed test.^c The **steps** argument allows further customization of the selection model, such as allowing for selection based on one-tailed rather than two-tailed hypothesis tests or allowing for selection that also favors “marginally significant” results with a two-tailed $p < 0.10$ to a lesser degree than “significant” results. However, in most scientific contexts, publication bias seems to conform well to one of the two specifications given in models (1) and (2) (Gelman & Carlin, 2014; Masicampo & Lalande, 2012; Mathur & VanderWeele, 2021; Wicherts, 2017).

In practice, we would recommend first fitting the more flexible model (2) to allow for the possibility of selection for both positive and negative “significant” results. If too few

^c One can optionally pass one’s own p -values to **weightfunct**, but it is important to note that doing so will override the package’s default calculation of p -values that are consistent with the specified selection mechanism.

studies fall into one of the three categories (positive and “significant”, negative and “significant”, and “nonsignificant”), the function will issue a warning, in which case we would recommend then fitting the simpler model (1) after coding point estimates’ signs such that the hypothesized effect direction is positive.

The `weightfunct` function reports the results of a standard random-effects meta-analysis without correction for publication bias, followed by analogous estimates after adjustment for publication bias. For example, for the meta-analysis on interracial dyads, the bias-corrected meta-analytic mean is $r = 0.04$ (95% CI: [-0.02, 0.09], $p = 0.087$). This p -value corresponds to a test of the pooled estimate versus an effect size of 0 in the bias-corrected model. The function also estimates “weights” for each interval of p -values defined by the cutoffs. For the meta-analysis on interracial dyads, for example, the estimated weight for the interval $0.025 < p_+ < 0.975$ is 0.19. Because *one*-tailed p -values in this interval correspond to *two*-tailed p -values greater than 0.05 and with estimates in either direction, this estimated weight indicates that “non-significant” two-tailed p -values are estimated to be only 19% as likely to be published compared to “significant” two-tailed p -values with point estimates in the expected positive direction. For the interval $0.975 < p_+ < 1$ (corresponding to two-tailed p -values less than 0.05 with negative estimates), the estimated weight is 1.05, indicating that studies that are “significant” but with point estimates in the unexpected negative direction are approximately as likely to be published as studies that are “significant” and positive. However, this estimate has considerable uncertainty because only 6 studies were “significant” and negative. Finally, the function provides a likelihood ratio test assessing for the presence of publication bias. We would suggest that meta-analysts characterize the evidence for publication bias in terms of the p -value of this test and the severity of publication bias in terms of the estimated weights themselves. Other software packages are available for fitting similar selection models (e.g., Rufibach, 2015; Viechtbauer, 2010).

Further Evidence That Funnel Plot Methods Alone Are Not Adequate

Simulation studies suggest that funnel plot methods can perform poorly at detecting and correcting for publication bias. The methods may spuriously detect publication bias when in fact there is none (Type I error) or, inversely, fail to detect publication bias when it does exist (Type II error) (e.g., Carter et al., 2019; Pustejovsky & Rodgers, 2019). These findings regarding Type I error are corroborated by a recent analysis of registered replication reports (RRRs), a publication type in which an article receives “in principle acceptance” based only on the introduction and methods sections (Maier et al., in press). In principle, then, RRRs should be subject to little (if any) selective reporting or publication bias (Chambers, 2013; Chambers et al., 2015). In an analysis of 28 RRRs, Egger’s regression found evidence for publication bias in 9 of 28 datasets, again suggesting a high false-positive rate. On the other hand, selection models did not find evidence for publication bias in any of these datasets (Maier et al., in press).

These findings regarding inflated Type I and Type II error rates in part reflect the statistical assumptions of funnel plot methods, as discussed above. The inflated error rates can also occur when effects are heterogeneous across studies; in these settings, many funnel plot methods are prone to detecting publication bias even when none exists (Egger et al., 1997; Higgins et al., 2019; Lau et al., 2006; Pustejovsky & Rodgers, 2019). However, in practice, meta-analyses often show moderate to high heterogeneity (McShane et al., 2016; Mathur and VanderWeele, 2021; Rhodes et al., 2015; van Erp et al., 2017), for example because the effect size differs across populations (Rothstein et al., 2005). Additionally, the standard errors of many effect-size measures (e.g., standardized mean differences) are arithmetically related to the effect size itself. This can induce an artifactual correlation between point estimates and their standard errors that funnel plot methods cannot distinguish from correlation induced by publication bias.^d

^d For effect-size measures that are artifactually correlated with their standard errors, one can sometimes use a modified standard error calculation that reduces artifactual correlation (Harbord et al., 2006; Peters

Discussion

Funnel plots and selection models can provide different insights in their assessments of publication bias, as we have illustrated using a previously published meta-analysis. This largely reflects the methods' differing assumptions: funnel plot methods assume that publication bias in fact operates based on effect sizes and standard errors and does not affect the largest studies, whereas selection models make the often more realistic assumption that publication bias operates on the "statistical significance" of p -values. Furthermore, funnel plot methods do not directly accommodate effect heterogeneity, whereas selection models do.

Nevertheless, selection models are not a panacea, nor are they the only reasonable methods to supplement funnel plots when assessing publication bias. Like all statistical methods to assess publication bias, selection models do require statistical assumptions. Most selection model specifications assume that, prior to selection due to publication bias, the true effects are normally distributed, independent, and not correlated with the point estimates' standard errors, assumptions that are also standard in random-effect meta-analysis more generally. Publication bias may not always conform exactly to the assumed p -value cutoffs, potentially compromising selection model estimates (Vevea & Hedges, 1995), but as we have discussed, empirical evidence suggests that the assumed model of publication bias often holds well in practice. Also, selection models perform best in large meta-analyses and might fail to converge or provide imprecise estimates if there are few studies in one or more of the p -value intervals (Carter et al., 2019; McShane et al., 2016; Terrin et al., 2003). In these settings, or when the normality or independence assumptions are considered implausible, it may be more informative to report sensitivity

et al., 2006; Pustejovsky & Rodgers, 2019; Rücker et al., 2008). Doing so can improve the performance of funnel plot methods when there is some heterogeneity, but some such methods may still perform poorly under moderate or high heterogeneity (Harbord et al., 2006; Peters et al., 2006), and simulations indicate that selection models still tend to perform better overall (Pustejovsky & Rodgers, 2019).

analyses that characterize how severe publication bias would hypothetically have to be in order to “explain away” the results of a meta-analysis rather than attempting to estimate and correct for the actual severity of publication bias (Mathur & VanderWeele, 2020). In addition, Bayesian selection models can mitigate estimation problems in cases where few primary studies are available (Larose & Dey, 1998; Smith et al., 2001). There are also Bayesian publication bias adjustment methods that allow simultaneous application of recent extensions to funnel plot methods and selection models (Bartoš et al., 2021; Maier et al., in press).

Shifting reporting practices away from exclusive focus on funnel plot methods and additionally (or alternatively) considering the results of selection models, as one reasonable alternative, would considerably improve our understanding of publication bias in meta-analyses. In providing concrete guidance on how to fit and interpret selection models using existing user-friendly software and in demonstrating that they can provide additional insights beyond those provided by funnel plot methods alone, we hope to open the door to more widespread adoption.

Funding

This research was supported by (1) Franklin E. Fetzner Memorial Trust (award R2020-16); (2) the NIH-funded Biostatistics, Epidemiology and Research Design (BERD) Shared Resource of Stanford University’s Clinical and Translational Education and Research (UL1TR003142); (3) the Biostatistics Shared Resource (BSR) of the NIH-funded Stanford Cancer Institute (P30CA124435); (4) the Quantitative Sciences Unit through the Stanford Diabetes Research Center (P30DK116074); and (5) NIH grant R01 CA222147.

Reproducibility

All data and code required to reproduce the applied examples is publicly available: <https://osf.io/37y9f/>.

References

- Bartoš, F., Maier, M., Wagenmakers, E.-J., Doucouliagos, H., & Stanley, T. D. (2021). No need to choose: Robust Bayesian meta-analysis with competing publication bias adjustment methods. <https://doi.org/10.31234/osf.io/kvsp7>
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons.
- Card, N. A. (2015). *Applied meta-analysis for social science research*. Guilford Publications.
- Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*, 2(2), 115–144. <https://doi.org/10.1177/2515245919847196>
- Chambers, C. D. (2013). Registered reports: A new publishing initiative at cortex. *Cortex*, 49(3), 609–610. <https://doi.org/10.1016/j.cortex.2012.12.016>
- Chambers, C. D., Dienes, Z., McIntosh, R. D., Rotshtein, P., & Willmes, K. (2015). Registered reports: Realigning incentives in scientific publishing. *Cortex*, 66, A1–A2.
- Coburn, K. M., Vevea, J. L., & Coburn, M. K. M. (2019). weightr: Estimating weight-function models for publication bias [R package version 2.0.2]. <https://CRAN.R-project.org/package=weightr>.
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455–463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109), 629–634. <https://doi.org/10.1136/bmj.315.7109.629>
- Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type s (sign) and type m (magnitude) errors. *Perspectives on Psychological Science*, 9(6), 641–651. <https://doi.org/10.1177/1745691614551642>

- Harbord, R. M., Egger, M., & Sterne, J. A. (2006). A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Statistics in Medicine*, 25(20), 3443–3457. <https://doi.org/10.1002/sim.2380>
- Hedges, L. V. (1984). Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational Statistics*, 9(1), 61–85. <https://doi.org/10.3102/10769986009001061>
- Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.
- Ioannidis, J. P., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials*, 4(3), 245–253. <https://doi.org/10.1177/1740774507079441>
- Iyengar, S., & Greenhouse, J. B. (1988). Selection models and the file drawer problem. *Statistical Science*, 3(1), 109–117. <https://doi.org/10.1214/ss/1177013012>
- Larose, D. T., & Dey, D. K. (1998). Modeling publication bias using weighted distributions in a Bayesian framework. *Computational Statistics & Data Analysis*, 26(3), 279–302. [https://doi.org/10.1016/S0167-9473\(97\)00039-X](https://doi.org/10.1016/S0167-9473(97)00039-X)
- Lau, J., Ioannidis, J. P., Terrin, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. *BMJ*, 333(7568), 597–600. <https://doi.org/10.1136/bmj.333.7568.597>
- Maier, M., Bartoš, F., & Wagenmakers, E.-J. (in press). Robust Bayesian meta-analysis: Addressing publication bias with model-averaging. <https://doi.org/10.31234/osf.io/u4cns>
- Masicampo, E., & Lalande, D. R. (2012). A peculiar prevalence of p values just below .05. *Quarterly Journal of Experimental Psychology*, 65(11), 2271–2279. <https://doi.org/10.1080/17470218.2012.711335>

- Mathur, M. B., & VanderWeele, T. J. (2020). Sensitivity analysis for publication bias in meta-analyses. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69(5), 1091–1119. <https://doi.org/10.1111/rssc.12440>
- Mathur, M. B., & VanderWeele, T. J. (2021). Estimating publication bias in meta-analyses of peer-reviewed studies: A meta-meta-analysis across disciplines and journal tiers. *Research Synthesis Methods*, 12(2), 176–191. <https://doi.org/10.1002/jrsm.1464>
- McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science*, 11(5), 730–749. <https://doi.org/10.1177/1745691616662243>
- McShane, B. B., & Gal, D. (2017). Statistical significance and the dichotomization of evidence. *Journal of the American Statistical Association*, 112(519), 885–895. <https://doi.org/10.1080/01621459.2017.1289846>
- Nelson, N., Rosenthal, R., & Rosnow, R. L. (1986). Interpretation of significance levels and effect sizes by psychological researchers. *American Psychologist*, 41(11), 1299.
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *Journal of the American Medical Association*, 295(6), 676–680. <https://doi.org/10.1001/jama.295.6.676>
- Pustejovsky, J. E., & Rodgers, M. A. (2019). Testing for funnel plot asymmetry of standardized mean differences. *Research Synthesis Methods*, 10(1), 57–71. <https://doi.org/10.1002/jrsm.1332>
- Rhodes, K. M., Turner, R. M., & Higgins, J. P. (2015). Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. *Journal of Clinical Epidemiology*, 68(1), 52–60. <https://doi.org/10.1016/j.jclinepi.2014.08.012>
- Rosenthal, R. (1986). Meta-analytic procedures for social science research. *Educational Researcher*, 15(8), 18–20. <https://doi.org/10.3102/0013189X015008018>

- Rosenthal, R., & Gaito, J. (1963). The interpretation of levels of significance by psychological researchers. *The Journal of Psychology*, 55(1), 33–38.
<https://doi.org/10.1080/00223980.1963.9916596>
- Rosenthal, R., & Gaito, J. (1964). Further evidence for the cliff effect in interpretation of levels of significance. *Psychological Reports*, 15(2), 570.
<https://doi.org/10.2466/pr0.1964.15.2.570>
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Wiley Online Library.
- Rücker, G., Schwarzer, G., & Carpenter, J. (2008). Arcsine test for publication bias in meta-analyses with binary outcomes. *Statistics in Medicine*, 27(5), 746–763.
<https://doi.org/10.1002/sim.2971>
- Rufibach, K. (2015). Package ‘selectMeta’ [R package].
<https://cran.r-project.org/web/packages/selectMeta/index.html>
- Smith, D. D., Givens, G. H., & Tweedie, R. L. (2001). Adjustment for publication bias and quality bias in Bayesian meta-analysis. *Biostatistics Basel*, 4, 277–304.
- Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, 5(1), 60–78.
<https://doi.org/10.1002/jrsm.1095>
- Stanley, T. D., & Doucouliagos, H. (2017). Neither fixed nor random: Weighted least squares meta-regression. *Research Synthesis Methods*, 8(1), 19–42.
<https://doi.org/10.1002/jrsm.1211>
- Stanley, T. D., Doucouliagos, H., & Ioannidis, J. P. (2017). Finding the power to reduce publication bias. *Statistics in Medicine*, 36(10), 1580–1598.
<https://doi.org/10.1002/sim.7228>
- Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., et al. (2011). Recommendations for

- examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, 343. <https://doi.org/10.1136/bmj.d4002>
- Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine*, 22(13), 2113–2126. <https://doi.org/10.1002/sim.1461>
- Toosi, N. R., Babbitt, L. G., Ambady, N., & Sommers, S. R. (2012). Dyadic interracial interactions: A meta-analysis. *Psychological bulletin*, 138(1), 1–17. <https://doi.org/10.1037/a0025767>
- van Erp, S., Verhagen, J., Grasman, R. P., & Wagenmakers, E.-J. (2017). Estimates of between-study heterogeneity for 705 meta-analyses reported in psychological bulletin from 1990–2013. *Journal of Open Psychology Data*, 5(1). <https://doi.org/10.5334/jopd.33>
- Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, 60(3), 419–435. <https://doi.org/10.1007/BF02294384>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <https://www.jstatsoft.org/v36/i03/>
- Wicherts, J. M. (2017). The weak spots in contemporary science (and how to fix them). *Animals*, 7(12), 90. <https://doi.org/10.3390/ani7120090>