Supplemental materials accompanying "Correcting for bias in psychology: A comparison of meta-analytic methods"

Evan C. Carter*
U.S. Army Research Laboratory, Aberdeen, MD, USA

Felix D. Schönbrodt*
Ludwig-Maximilians-University, Munich, Germany

Will M. Gervais
University of Kentucky, Lexington, KY, USA

Joseph Hilgard
University of Pennsylvania, Philadelphia, PA, USA

This is supplementary material for:

Carter, E. C., Schönbrodt, F. D., Gervais, W., & Hilgard, J. (2017). Correcting for bias in psychology: A comparison of meta-analytic methods. Retrieved from https://osf.io/rf3ys

Supplemental methods

Effect size estimates

Each observation in a meta-analytic data set must include, at a minimum, an estimate of the effect size and an estimate of the variance of that effect size estimate. Because meta-analyses are usually applied to studies with dependent variables measured on different scales, effect size estimates are typically standardized. To synthesize the observed studies, one would need to first transform the results of each study into an effect size measure such as the standardized mean difference, or Cohen’s $d$, given as

$$d = \frac{M_1 - M_2}{S},$$

where $M_1$ and $M_2$ are the means of the two groups and $S$ is the pooled standard error of the two groups. $S$ can be calculated as

$$S = \sqrt{\frac{(n_1 - 1)v_1 + (n_2 - 1)v_2}{(n_1 + n_2 - 2)}},$$

Correspondence concerning this article should be addressed to Evan Carter, Email: evan.c.carter@gmail.com. *These authors contributed equally to this work.
where \( n \) and \( v \) are the sample sizes and variances of the groups. For Cohen’s \( d \), the variance can be calculated as

\[
v_d = \frac{n_1 + n_2}{n_2n_2} + \frac{d^2}{2(n_1 + n_2 - 2)} \cdot \frac{n_1 + n_2}{n_2n_2 - 2}.
\]

We calculated effect size estimates at the study-level using the above formulas.

**Funnel plots**

The influence of bias in meta-analysis can sometimes be seen by comparing the effect size estimates to the standard errors of those estimates (or some other indicator of sample size) with a funnel plot (Light & Pillemer, 1984). In a typical funnel plot, the reported effect size is plotted on the x-axis and the standard error is plotted on the inverted y-axis. The most precise estimates (i.e., those with the smallest standard error and largest sample) will tend to converge on the true effect size, whereas the more imprecise estimates will spread evenly on either side of the true effect, with studies equally likely to overestimate or underestimate the true effect. That is, the amount of deviation from the true effect increases as estimates become more imprecise, leading to a funnel-like pattern (Figure 1A). In the presence of bias, fewer studies will be present in the lower corner of the funnel where results would be non-significant or of the wrong sign (Figure 1B). In this case, the funnel plot will appear asymmetrical, with more imprecise studies finding larger effects than more precise studies. The blue triangle displays the region of non-significant studies. In the case of complete publication bias (i.e., only significant studies entered the published literature), no studies are present in the non-significant region. In this way, a funnel plot can reveal patterns that may indicate bias.

Figure 1A shows funnel plots of simulated meta-analytic data sets. These data sets vary in the true values of the underlying effect, \( \delta \), and heterogeneity, given as the standard deviation of the distribution of true effects, \( \tau \). Note that in this panel, none of these meta-analyses have been affected by bias, and the difference between the random-effects model estimate (marked as a solid vertical line and a “X” along the horizontal axis) and the true value (marked as a dashed vertical line and a filled dot) is very close to zero. Figure 1B, in contrast, shows newly generated samples from these same conditions but under complete publication bias. Note the clear rightward asymmetry of the funnel plots in Figure 1B as compared to in Figure 1A, as well as the resulting overestimation in the random-effects model estimates: Along the horizontal axis, each X has been shifted to the right of the true value.

As is illustrated in Figure 1B, publication bias induces a relationship—in this case, a positive correlation—between effect size estimates and their standard errors. However, such a correlation can also have benign causes. It may be, for example, that expensive, small-sample manipulations have stronger effects than inexpensive, large-sample manipulations. Similarly, when a literature contains both large and small effects, and researchers use power analyses to plan their samples sizes accordingly, the large effects will be studied with smaller, less-precise samples. Sequential designs can also induce this correlation (Lakens, 2014; Schönbrodt, Wagenmakers, Zehetleitner, & Perugini, 2015): Studies measuring a large effect can stop early for efficacy, whereas studies measuring a small effect can stop at later stages of the sequential design after continuing data collection. Finally, sometimes a relationship
between effect size and standard error is built into the calculation of an effect size’s precision (e.g., see the equation for Cohen’s $d$ above, which includes sample size, as well as Macaskill, Walter, & Irwig, 2001; Peters, Sutton, Jones, Abrams, & Rushton, 2006). In these scenarios, effect size and standard error would also be correlated, but not because of bias.

Because this correlation between sample size and effect size can have several causes, some bias-inducing and others benign, such a correlation is typically called a “small-study effect” (Sterne, Gavaghan, & Egger, 2000). Such small-study effects do not necessarily indicate publication bias or QRPs. Because several of the methods we examine here adjust for small-study effects generally as though such effects always represent publication bias, they may overadjust under certain conditions.

**Estimation of the Monte Carlo simulation error**

To check whether 1000 simulations were enough to get sufficiently stable estimates, we ran simulations with 10,000 replications for a random selection of 30 conditions. Then, these 10,000 runs were divided in 10 batches of 1000 simulations each. Monte Carlo simulation error, which is the standard deviation of the Monte Carlo estimator taken across repetitions of the simulation (Koehler, Brown, & Haneuse, 2009), was on average 0.002 for the effect size estimate (95% quantile: 0.008), and on average 0.003 (95% quantile: 0.008) for the RMSE estimate.

**Percentages of Valid Estimates**

Some methods do not always return a valid estimate, either due to inherent limitations (e.g., $p$-curve and $p$-uniform only work if $\geq 1$ significant studies are in the set), or due to failed convergence of the estimation method. All reported results have to be read as conditional results: In all cases where an estimation method produced an estimate, the method had the reported performance.

Random effects, PET-PEESE, and WAAP-WLS always returned an estimate, but not the other methods. Table 1 shows the rates of valid estimates in each condition.

**Technical notes: PET, PEESE, PET-PEESE**

The typical effect of PET and PEESE in the presence of publication bias is a downward correction of the meta-analytic estimate. When no publication bias is present, however, it can happen that random variations in the sample induce a negative correlation between sample size and effect size, which leads to an upward correction. In the current simulations we kept these upward corrections. In an applied setting, however, we recommend that analysts be very skeptical when PET or PEESE have a slope of the reversed sign.

Notably, both PET and PEESE are examples of weighted-least squares meta-regression and are therefore distinct in some ways from the fixed- and random-effects meta-analysis models described above. The specifics of this difference are discussed in detail elsewhere (Thompson & Sharp, 1999; Stanley & Doucouliagos, 2015); however, in practice, the result of the difference is that the estimates from weighted-least squares meta-regression models will have relatively larger standard errors, and thus, relatively wider confidence intervals than standard meta-analysis models. This is not necessarily a negative in the face
of heterogeneity and publication bias, and authors have argued for the use of both types of models (Thompson & Sharp, 1999; Stanley & Doucouliagos, 2015; Moreno et al., 2012).

We estimated PET by the formula \( \text{PET.lm} \leftarrow \text{lm}(d \sim \sqrt{v}, \text{weights}=1/v) \). An alternative approach is to use an additive error term, which is akin to the standard meta-analysis: \( \text{PET.rma} \leftarrow \text{rma}(y_i = d, v_i = v, \text{mods}=\sqrt{v}, \text{method}="\text{REML}) \). The same holds true for PEESE, but replacing \( \sqrt{v} \) with \( v \).

It is worth noting that the downward bias shown by PET and PEESE when \( \delta \) was large may be due to the fact that the standard error of \( d \) is a function of both the sample size and the observed \( d \) (see the equation in the introduction). This relationship between \( d \) and its standard error is such that larger \( d \) leads to larger standard errors, thereby creating a small-study effect that mimics that of publication bias and leads to an overadjustment.

**Technical notes: 3PSM**

We employed the `weightfunct` function provided by the `weightr` package (Coburn & Vevea, 2017).

**Technical notes: p-uniform**

We used the default “P” method from the `puniform` package (van Aert, 2017) for R, which relies on the Irwin-Hall distribution. The package returns a one-tailed \( p \)-value by default. As all other methods use two-tailed \( p \)-values, we doubled the resulting \( p \)-value to achieve comparable results.

**Technical notes: p-curve**

We used the test for right skew with the Stouffer method for the hypothesis test of evidential value. A non-significant skew means that \( H_0: \delta = 0 \) is not rejected. \( p \)-curve also provides a test that tests whether the \( p \)-curve is flatter than a reference line at 33% power, which can be seen as an indicator for lack of evidence and non-rejection of \( H_0 \). Typically both tests agree in the sense that when one test is significant, the other is not. In a small fraction of all simulations, however, both tests were significant, indicating that the \( p \)-curve is both flatter than the reference line and steeper than zero. By only considering the skewness test, these cases were treated as \( H_0 \) rejections.

**Assessing p-curve’s performance to recover the average effect size of all included studies**

The standard goal of a random effects meta-analysis is to estimate the mean and the variance of all conducted studies. Simonsohn, Nelson, and Simmons (2014), in contrast, put forward a different interpretation of the \( p \)-curve estimate: “It is the average effect size one expects to get if one were to rerun all studies included in the \( p \)-curve” (p. 667). That means, it does not attempt to recover the mean of all conducted studies, but rather the true mean of all significant studies (i.e., the studies which are entered into the \( p \)-curve analysis), corrected for publication bias.

In our main analyses we evaluated all methods with regard to their ability to recover the mean of all studies. In this supplemental analysis, we report how well \( p \)-curve estimates
the mean of studies submitted to the p-curve analysis. Figure 2 shows the mean error of the p-curve effect size estimate relative to the mean of the true study-level effect sizes that have been submitted to the analysis. If no QRPs are present, p-curve perfectly recovers that quantity, regardless of the level of publication bias and δ. The only deviations can be seen at δ = 0, but this might be attributed to the fact that only very few (mostly 1 or 2) significant studies are submitted to the p-curve analysis in these conditions. The combination of QRPs as we have modeled them induced a downward bias, which was stronger for smaller δs and for less heterogeneity. These results are consistent with previous simulation results (Simonsohn et al., 2014). Note, however, that in a simulation from van Aert, Wicherts, and van Assen (2016), some other QRPs lead to an overestimation of p-curve and p-uniform estimates. These overestimates, however, have been observed relative to the mean of all conducted studies.

Results for medium publication bias with no QRPs

**Type I error rate.** Publication bias led to a sharp increase in Type I error rates in random-effects meta-analysis. At k = 10, the false positive rate was 51%; at k = 60, the false positive rate was 100%. Trim-and-fill had noticeably lower, but still elevated, Type I error rates (34% at k = 10; 62% at k = 60). WAAP-WLS similarly improved on random effects, but error rates remained high (38% at k = 10; 98% at k = 60). 3PSM had lower Type I error rates, but error rates increased with k (19% at k = 10; 63% at k = 60). p-curve and PET-PEESE had approximately conservative Type I error rates (4–6%), while p-uniform was exceedingly conservative (2–3%).

The addition of heterogeneity led to a slight increase in the Type I error rate of random-effects analysis. This was noticeable only when k = 10 (59%); at k = 60, Type I error rates remained at 100%. Heterogeneity also caused a slight increase in Type I error rates of trim-and-fill (+~12pp). For WAAP-WLS, heterogeneity increased Type I error at k = 10 (+7pp) but reduced errors at k = 60 (-46pp). Heterogeneity also increased the error rates of PET-PEESE (32% at k = 10; 48% at k = 60) as well as p-curve and p-uniform (25% at k = 10; 80% at k = 60). Heterogeneity increased 3PSM’s Type I error at low k (+3pp) but reduced it at high k (-27pp).

**Power.** Power to detect δ = 0.5 with random-effects meta-analysis was 100%. All methods had good (85%+: PET-PEESE, p-uniform) or great (90%+: WAAP-WLS, trim-and-fill, p-curve, 3PSM) power.

Heterogeneity did not affect the power of random-effects meta-analysis. When k was small, heterogeneity caused a slight loss of power in 3PSM (90%) and in WAAP-WLS (83%); it caused a noticeable loss of power in PET-PEESE (76% at k = 10, a 12pp drop). Other methods were less affected, and all methods had 99%+ power at k = 60.

**ME.** Under the null, medium publication bias lead to a slight bias in estimates; random-effects meta-analysis estimated the null effect as 0.16. Trim and fill reduced this bias and benefited from increasing k (ME = 0.11 at k = 10; ME = 0.08 at k = 60). WAAP-WLS similarly reduced the bias, but not by as much as trim-and-fill (ME = 0.13, 0.11). 3PSM, p-curve, and p-uniform all showed upward bias at k = 10, with 3PSM having the least bias (ME = 0.13), but at k = 60 all exhibited similar moderate upward bias (~.09). Only PET-PEESE was unbiased.
When the null was false, random-effects still slightly overestimated the true effect ($\sim 0.05$). Bias in all adjustments was minimal; trim-and-fill had bias $\leq 0.02$, WAAP-WLS had bias $\leq 0.01$, and $p$-curve, $p$-uniform, and 3PSM were all unbiased. PET-PEESE, however, tended to underestimate the true effect slightly ($\text{ME} = -0.04$ at $k = 10$; $\text{ME} = -0.02$ at $k = 60$).

Adding heterogeneity to a true effect of zero tended to create further upward bias. Random-effects meta-analysis ($\sim 0.23$), trim-and-fill ($\sim 0.15$), and WAAP-WLS ($\sim 0.15$) all showed moderate upward bias. PET-PEESE developed a slight upward bias (0.06). Heterogeneity caused upward bias in $p$-curve and $p$-uniform ($\sim 0.25$). Heterogeneity had little effect on 3PSM’s moderate upward bias.

Adding heterogeneity to a true nonzero effect also tended to create upward bias. Random-effects, trim-and-fill, and WAAP-WLS gained an additional small amount of bias (increases of about 0.07, 0.03, and 0.03, respectively). $p$-curve and $p$-uniform also demonstrated slight upward bias ($\sim 0.07$), as did 3PSM ($\sim 0.02$). PET-PEESE’s slight downward bias was very slightly reduced by heterogeneity.

**RMSE.** Given the null, publication bias caused an increase in RMSE. With only 10 studies, most adjustments failed to improve RMSE; trim-and-fill and WAAP-WLS caused slight benefits in RMSE ($\sim -0.03$). PET-PEESE did not change RMSE, 3PSM slightly increased RMSE (+0.03), $p$-uniform increased RMSE substantially (+0.17), and $p$-curve increased RMSE dramatically (+0.77). With 60 studies, PET-PEESE substantially reduced RMSE; trim-and-fill, 3PSM, and WAAP-WLS reduced RMSE slightly; and $p$-curve and $p$-uniform very slightly increased RMSE.

Given a true effect, the increase in RMSE due to publication bias was very slight (0.02). At $k = 10$, trim-and-fill and WAAP-WLS reduce RMSE by .02, 3PSM reduces RMSE by .01, $p$-curve and $p$-uniform increase RMSE by .02, and PET-PEESE increases RMSE by .04. At $k = 60$, all adjustments slightly reduced RMSE (0.01–0.02).

Adding heterogeneity to a true effect of zero lead to a further increase in the RMSE of random-effects meta-analysis. At $k = 10$, trim-and-fill, WAAP-WLS, and 3PSM all yielded small improvements in RMSE ($\sim -0.04$). PET-PEESE did not affect RMSE, and $p$-curve and $p$-uniform increased RMSE (+0.60 with $p$-curve, +0.05 with $p$-uniform). At $k = 60$, most adjustments yielded benefits in RMSE, including 3PSM (-0.12), PET-PEESE (-0.09), trim-and-fill (-0.07), and WAAP-WLS (-0.07). By contrast, $p$-curve and $p$-uniform caused small increases in RMSE (+0.03), likely owing to their upward bias under heterogeneity.

Adding heterogeneity to a true nonzero effect lead to a further increase in the RMSE of random-effects meta-analysis. At small $k$, all methods except PET-PEESE lead to a small (0.01–0.03) decrease in RMSE; PET-PEESE increased RMSE noticeably (+0.07). At large $k$, all adjustments yielded a noticeable improvement in RMSE (0.03–0.06), with $p$-uniform giving the smallest benefit and 3PSM the largest benefit.

**95% CI coverage.** Given the null, even medium publication bias was enough to substantially harm 95% CI coverage, particularly as $k$ increased. All methods lead to some improvement in coverage. Benefits of WAAP-WLS were small. Trim-and-fill and 3PSM had slightly better coverage, particularly at higher $k$. Only PET-PEESE and $p$-uniform had appropriate coverage rates (94–95%).

Given a true effect, medium publication bias reduced random-effects CI coverage moderately, and coverage worsened as $k$ increased. In this case, trim-and-fill, 3PSM, WAAP-
WLS, and $p$-uniform were all able to establish appropriate (93–96%) coverage rates. PET-PEESE, however, suffered from some undercoverage ($\sim 87\%$), although this was still an improvement over random-effects meta-analysis.

Adding heterogeneity to a true effect of zero, coverage fell even further for random-effects meta-analysis. Trim-and-fill provided a modest benefit. Benefits of WAAP-WLS, PET-PEESE, and 3PSM were greater, with coverages of 50–75%. $p$-uniform improved coverage substantially at $k = 10$ (65%) but only modestly at $k = 60$ (16%).

Adding heterogeneity to a true nonzero effect also reduced CI coverage for random-effects meta-analysis. All adjustments yielded benefits in coverage. The greatest coverage was observed using $p$-uniform or WAAP-WLS with $k = 10$ (85%) and using 3PSM when $k = 60$ (90%).

**Detailed results for all conditions**

Figures 3 to 18 show mean error (ME), root mean square error (RMSE), rejection rates, and coverage probabilities for all methods and all conditions.

**References**

Coburn, K. M. & Vevea, J. L. (2017). *Weightr: Estimating weight-function models for publication bias*. R package version 1.1.2. Retrieved from https://CRAN.R-project.org/package=weightr

Koehler, E., Brown, E., & Haneuse, S. J.-P. A. (2009). On the assessment of monte carlo error in simulation-based statistical analyses. *The American Statistician, 63*(2), 155–162. doi:10.1198/tast.2009.0030

Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology, 44*(7), 701–710.

Light, R. J. & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Harvard University Press.

Macaskill, P., Walter, S. D., & Irwig, L. (2001). A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine, 20*(4), 641–654.

Moreno, S. G., Sutton, A. J., Thompson, J. R., Abrams, K. R., & Cooper, N. J. (2012). A generalized weighting regression-derived meta-analysis estimator robust to small-study effects and heterogeneity. *Statistics in Medicine, 31*(14), 1407–1417.

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. *JAMA, 295*(6), 676–680.

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2015). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science, 9*(6), 666–681.

Stanley, T. & Doucouliagos, H. (2015). Neither fixed nor random: Weighted least squares meta-analysis. *Statistics in Medicine, 34*(13), 2116–2127.
Sterne, J. A., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of clinical Epidemiology, 53*(11), 1119–1129.

Thompson, S. G. & Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine, 18*(20), 2693–2708.

van Aert, R. C. (2017). *Puniform: Meta-analysis methods correcting for publication bias*. R package version 0.0.3.

van Aert, R. C., Wicherts, J. M., & van Assen, M. A. (2016). Conducting meta-analyses based on p values: Reservations and recommendations for applying p-uniform and p-curve. *Perspectives on Psychological Science, 11*(5), 713–729.

| Abbreviation | Meaning |
|--------------|---------|
| RE           | Random-effects meta-analysis |
| TF           | Trim-and-fill |
| PT           | Precision effect test (PET) |
| PE           | Precision effect estimate with standard error (PEESE) |
| PP           | PET-PEESE |
| PC           | p-curve |
| PU           | p-uniform |
| 3P           | Three parameter selection model |
| qrpEnv       | QRP Environment (see main text) |
**Figure 1.** Funnel plots compare the effect size estimate against its standard error across studies. (A) In the absence of publication bias, data points form a symmetrical funnel, conforming closely to the true effect size when the standard error is small and spreading evenly when the standard error is large. Heterogeneity in the true effect size leads to greater spread. (B) Publication bias selectively removes non-statistically-significant effect size estimates. This censorship of small effect size estimates leads to an asymmetrical funnel and overestimation of the true effect size. Figure available at https://osf.io/rf3ys, under a CC-BY4.0 license.
| 10 | 0.0 | high | high | 0.2 | 95% | 100% | 100% | 96% |
| 10 | 0.0 | high | high | 0.4 | 95% | 100% | 100% | 90% |
| 10 | 0.2 | none | none | 0.0 | 98% | 88% | 88% | 91% |
| 10 | 0.2 | none | none | 0.2 | 98% | 91% | 91% | 95% |
| 10 | 0.2 | none | med | 0.0 | 99% | 100% | 100% | 84% |
| 10 | 0.2 | none | med | 0.2 | 98% | 100% | 100% | 95% |
| 10 | 0.2 | none | high | 0.0 | 98% | 100% | 100% | 70% |
| 10 | 0.2 | none | high | 0.2 | 98% | 100% | 100% | 94% |
| 10 | 0.2 | none | high | 0.4 | 99% | 100% | 100% | 90% |
| 10 | 0.2 | med | none | 0.0 | 99% | 99% | 99% | 94% |
| 10 | 0.2 | med | none | 0.2 | 100% | 100% | 100% | 98% |
| 10 | 0.2 | med | med | 0.0 | 95% | 100% | 100% | 91% |
| 10 | 0.2 | med | med | 0.2 | 95% | 100% | 100% | 96% |
| 10 | 0.2 | med | high | 0.0 | 93% | 100% | 100% | 89% |
| 10 | 0.2 | med | high | 0.2 | 94% | 100% | 100% | 95% |
| 10 | 0.2 | med | high | 0.4 | 96% | 100% | 100% | 98% |
| 10 | 0.2 | high | none | 0.0 | 99% | 100% | 100% | 94% |
| 10 | 0.2 | high | none | 0.2 | 100% | 100% | 100% | 99% |
| 10 | 0.2 | high | none | 0.4 | 100% | 100% | 100% | 100% |
| 10 | 0.2 | high | med | 0.0 | 98% | 100% | 100% | 94% |
| 10 | 0.2 | high | med | 0.2 | 95% | 100% | 100% | 95% |
| 10 | 0.2 | high | high | 0.0 | 92% | 100% | 100% | 90% |
| 10 | 0.2 | high | high | 0.2 | 92% | 100% | 100% | 95% |
| 10 | 0.2 | high | high | 0.4 | 95% | 100% | 100% | 99% |
| 10 | 0.5 | none | none | 0.0 | 99% | 100% | 100% | 96% |
| 10 | 0.5 | none | none | 0.2 | 98% | 100% | 100% | 98% |
| 10 | 0.5 | none | none | 0.4 | 100% | 100% | 100% | 100% |
| 10 | 0.5 | none | med | 0.0 | 98% | 100% | 100% | 94% |
| 10 | 0.5 | none | med | 0.2 | 98% | 100% | 100% | 99% |
| 10 | 0.5 | none | med | 0.4 | 100% | 100% | 100% | 100% |
| 10 | 0.5 | none | high | 0.0 | 97% | 100% | 100% | 93% |
| 10 | 0.5 | none | high | 0.2 | 96% | 100% | 100% | 97% |
| 10 | 0.5 | none | high | 0.4 | 98% | 100% | 100% | 99% |
| 10 | 0.5 | med | none | 0.0 | 98% | 100% | 100% | 95% |
| 10 | 0.5 | med | none | 0.2 | 98% | 100% | 100% | 98% |
| 10 | 0.5 | med | none | 0.4 | 100% | 100% | 100% | 99% |
| 10 | 0.5 | med | med | 0.0 | 96% | 100% | 100% | 93% |
| 10 | 0.5 | med | med | 0.2 | 94% | 100% | 100% | 97% |
| 10 | 0.5 | med | med | 0.4 | 98% | 100% | 100% | 99% |
| 10 | 0.5 | med | high | 0.0 | 95% | 100% | 100% | 94% |
| 10 | 0.5 | med | high | 0.2 | 93% | 100% | 100% | 97% |
| 10 | 0.5 | med | high | 0.4 | 99% | 100% | 100% | 99% |
| 10 | 0.5 | high | none | 0.0 | 97% | 100% | 100% | 95% |
| 10 | 0.5 | high | none | 0.2 | 97% | 100% | 100% | 99% |
| 10 | 0.5 | high | none | 0.4 | 99% | 100% | 100% | 100% |
| 10 | 0.5 | high | med | 0.0 | 96% | 100% | 100% | 94% |
| 10 | 0.5 | high | med | 0.2 | 95% | 100% | 100% | 98% |
| 10 | 0.5 | high | med | 0.4 | 98% | 100% | 100% | 99% |
| 10 | 0.5 | high | high | 0.0 | 95% | 100% | 100% | 93% |
| 10 | 0.5 | high | high | 0.2 | 92% | 100% | 100% | 96% |
| 10 | 0.5 | high | high | 0.4 | 96% | 100% | 100% | 100% |
| 10 | 0.8 | none | none | 0.0 | 98% | 100% | 100% | 95% |
| 10 | 0.8 | none | none | 0.2 | 99% | 100% | 100% | 98% |
| 10 | 0.8 | none | none | 0.4 | 100% | 100% | 100% | 99% |
| 10 | 0.8 | none | med | 0.0 | 98% | 100% | 100% | 94% |
| 10 | 0.8 | none | med | 0.2 | 97% | 100% | 100% | 98% |
| 10 | 0.8 | none | med | 0.4 | 99% | 100% | 100% | 100% |
| 10 | 0.8 | none | high | 0.0 | 97% | 100% | 100% | 93% |
| 10 | 0.8 | none | high | 0.2 | 97% | 100% | 100% | 97% |
| x  | y  | A   | B   | C   | D   |
|----|----|-----|-----|-----|-----|
| 10 | 0.8| none | high| 0.4 | 98% | 100%| 100%| 99% |
| 10 | 0.8| med  | none | 0.0 | 98% | 100%| 100%| 95% |
| 10 | 0.8| med  | none | 0.2 | 97% | 100%| 100%| 99% |
| 10 | 0.8| med  | none | 0.4 | 99% | 100%| 100%| 100%|
| 10 | 0.8| med  | med  | 0.0 | 98% | 100%| 100%| 94% |
| 10 | 0.8| med  | high | 0.2 | 97% | 100%| 100%| 98% |
| 10 | 0.8| med  | high | 0.4 | 98% | 100%| 100%| 99% |
| 10 | 0.8| high | none | 0.0 | 96% | 100%| 100%| 95% |
| 10 | 0.8| high | none | 0.2 | 97% | 100%| 100%| 98% |
| 10 | 0.8| high | none | 0.4 | 99% | 100%| 100%| 100%|
| 10 | 0.8| high | med  | 0.0 | 96% | 100%| 100%| 96% |
| 10 | 0.8| high | med  | 0.2 | 95% | 100%| 100%| 98% |
| 10 | 0.8| high | med  | 0.4 | 98% | 100%| 100%| 100%|
| 10 | 0.8| high | high | 0.0 | 94% | 100%| 100%| 96% |
| 10 | 0.8| high | high | 0.2 | 95% | 100%| 100%| 97% |
| 10 | 0.8| high | high | 0.4 | 98% | 100%| 100%| 99% |
| 30 | 0.0| none | none | 0.0 | 98% | 53% | 53% | 76% |
| 30 | 0.0| none | none | 0.2 | 99% | 93% | 93% | 97% |
| 30 | 0.0| none | none | 0.4 | 100%| 100%| 100%| 100%|
| 30 | 0.0| none | med  | 0.0 | 94% | 100%| 100%| 89% |
| 30 | 0.0| none | med  | 0.2 | 99% | 100%| 100%| 90% |
| 30 | 0.0| none | med  | 0.4 | 100%| 100%| 100%| 100%|
| 30 | 0.0| high | none | 0.0 | 91% | 90% | 90% | 96% |
| 30 | 0.0| high | none | 0.2 | 99% | 100%| 100%| 98% |
| 30 | 0.0| high | none | 0.4 | 100%| 100%| 100%| 100%|
| 30 | 0.0| high | med  | 0.0 | 98% | 100%| 100%| 86% |
| 30 | 0.0| high | med  | 0.2 | 99% | 100%| 100%| 90% |
| 30 | 0.0| high | med  | 0.4 | 100%| 100%| 100%| 100%|
| 30 | 0.0| high | high | 0.0 | 85% | 100%| 100%| 88% |
| 30 | 0.0| high | high | 0.2 | 92% | 100%| 100%| 95% |
| 30 | 0.0| high | high | 0.4 | 98% | 100%| 100%| 100%|
| 30 | 0.0| high | high | 0.6 | 80% | 100%| 100%| 90% |
| 30 | 0.0| high | high | 0.8 | 90% | 100%| 100%| 95% |
| 30 | 0.0| high | high | 1.0 | 96% | 100%| 100%| 100%|
| 30 | 0.2| none | none | 0.0 | 98% | 100%| 100%| 86% |
| 30 | 0.2| none | none | 0.2 | 99% | 100%| 100%| 90% |
| 30 | 0.2| none | none | 0.4 | 100%| 100%| 100%| 90% |
| 30 | 0.2| none | med  | 0.0 | 95% | 90% | 90% | 90% |
| 30 | 0.2| none | med  | 0.2 | 99% | 100%| 100%| 98% |
| 30 | 0.2| none | med  | 0.4 | 100%| 100%| 100%| 100%|
| 30 | 0.2| none | high | 0.0 | 92% | 100%| 100%| 94% |
| 30 | 0.2| none | high | 0.2 | 96% | 100%| 100%| 98% |
| 30 | 0.2| none | high | 0.4 | 98% | 100%| 100%| 100%|
| 30 | 0.2| med  | none | 0.0 | 99% | 90% | 90% | 90% |
| 30 | 0.2| med  | none | 0.2 | 98% | 100%| 100%| 90% |
| 30 | 0.2| med  | none | 0.4 | 100%| 100%| 100%| 100%|
| 30 | 0.2| med  | med  | 0.0 | 98% | 100%| 100%| 95% |
| 30 | 0.2| med  | med  | 0.2 | 99% | 100%| 100%| 99% |
| 30 | 0.2| med  | med  | 0.4 | 100%| 100%| 100%| 100%|
| 30 | 0.2| med  | high | 0.0 | 87% | 90% | 90% | 90% |
| 30 | 0.2| med  | high | 0.2 | 82% | 100%| 100%| 94% |
| 30 | 0.2| med  | high | 0.4 | 91% | 100%| 100%| 100%|
|     |     |     |     |     |
|-----|-----|-----|-----|-----|
| 30  | 0.2 | high  | none | 0.0  | 99% | 100% | 100% | 87% |
| 30  | 0.2 | high  | none | 0.2  | 100%| 100% | 100% | 99% |
| 30  | 0.2 | high  | none | 0.4  | 100%| 100% | 100% | 96% |
| 30  | 0.2 | high  | med  | 0.0  | 86% | 100% | 100% | 79% |
| 30  | 0.2 | high  | med  | 0.2  | 90% | 100% | 100% | 96% |
| 30  | 0.2 | high  | med  | 0.4  | 98% | 100% | 100% | 100% |
| 30  | 0.2 | high  | high | 0.0  | 84% | 100% | 100% | 74% |
| 30  | 0.2 | high  | high | 0.2  | 78% | 100% | 100% | 95% |
| 30  | 0.2 | high  | high | 0.4  | 89% | 100% | 100% | 100% |
| 30  | 0.5 | none  | none | 0.0  | 98% | 100% | 100% | 87% |
| 30  | 0.5 | none  | none | 0.2  | 98% | 100% | 100% | 98% |
| 30  | 0.5 | none  | none | 0.4  | 100%| 100% | 100% | 100% |
| 30  | 0.5 | none  | med  | 0.0  | 95% | 100% | 100% | 78% |
| 30  | 0.5 | none  | med  | 0.2  | 95% | 100% | 100% | 99% |
| 30  | 0.5 | none  | med  | 0.4  | 99% | 100% | 100% | 100% |
| 30  | 0.5 | med   | none | 0.0  | 93% | 100% | 100% | 78% |
| 30  | 0.5 | med   | none | 0.2  | 91% | 100% | 100% | 98% |
| 30  | 0.5 | med   | none | 0.4  | 98% | 100% | 100% | 100% |
| 30  | 0.5 | med   | none | 0.0  | 94% | 100% | 100% | 86% |
| 30  | 0.5 | med   | none | 0.2  | 97% | 100% | 100% | 98% |
| 30  | 0.5 | med   | none | 0.4  | 100%| 100% | 100% | 100% |
| 30  | 0.5 | med   | med  | 0.0  | 88% | 100% | 100% | 82% |
| 30  | 0.5 | med   | med  | 0.2  | 83% | 100% | 100% | 98% |
| 30  | 0.5 | med   | med  | 0.4  | 96% | 100% | 100% | 100% |
| 30  | 0.5 | med   | high | 0.0  | 91% | 100% | 100% | 83% |
| 30  | 0.5 | med   | high | 0.2  | 79% | 100% | 100% | 98% |
| 30  | 0.5 | med   | high | 0.4  | 91% | 100% | 100% | 100% |
| 30  | 0.5 | high  | none | 0.0  | 90% | 100% | 100% | 87% |
| 30  | 0.5 | high  | none | 0.2  | 95% | 100% | 100% | 99% |
| 30  | 0.5 | high  | none | 0.4  | 100%| 100% | 100% | 100% |
| 30  | 0.5 | high  | med  | 0.0  | 94% | 100% | 100% | 84% |
| 30  | 0.5 | high  | med  | 0.2  | 82% | 100% | 100% | 96% |
| 30  | 0.5 | high  | med  | 0.4  | 93% | 100% | 100% | 100% |
| 30  | 0.5 | high  | high | 0.0  | 90% | 100% | 100% | 86% |
| 30  | 0.5 | high  | high | 0.2  | 74% | 100% | 100% | 98% |
| 30  | 0.5 | high  | high | 0.4  | 88% | 100% | 100% | 100% |
| 30  | 0.8 | none  | none | 0.0  | 97% | 100% | 100% | 86% |
| 30  | 0.8 | none  | none | 0.2  | 98% | 100% | 100% | 99% |
| 30  | 0.8 | none  | none | 0.4  | 100%| 100% | 100% | 100% |
| 30  | 0.8 | none  | med  | 0.0  | 95% | 100% | 100% | 87% |
| 30  | 0.8 | none  | med  | 0.2  | 94% | 100% | 100% | 100% |
| 30  | 0.8 | none  | med  | 0.4  | 99% | 100% | 100% | 100% |
| 30  | 0.8 | none  | high | 0.0  | 94% | 100% | 100% | 85% |
| 30  | 0.8 | none  | high | 0.2  | 93% | 100% | 100% | 99% |
| 30  | 0.8 | none  | high | 0.4  | 98% | 100% | 100% | 100% |
| 30  | 0.8 | med   | none | 0.0  | 96% | 100% | 100% | 88% |
| 30  | 0.8 | med   | none | 0.2  | 93% | 100% | 100% | 98% |
| 30  | 0.8 | med   | none | 0.4  | 100%| 100% | 100% | 100% |
| 30  | 0.8 | med   | med  | 0.0  | 93% | 100% | 100% | 88% |
| 30  | 0.8 | med   | med  | 0.2  | 87% | 100% | 100% | 99% |
| 30  | 0.8 | med   | med  | 0.4  | 97% | 100% | 100% | 100% |
| 30  | 0.8 | med   | high | 0.0  | 93% | 100% | 100% | 87% |
| 30  | 0.8 | med   | high | 0.2  | 88% | 100% | 100% | 99% |
| 30  | 0.8 | med   | high | 0.4  | 96% | 100% | 100% | 100% |
| 30  | 0.8 | high  | none | 0.0  | 91% | 100% | 100% | 85% |
| 30  | 0.8 | high  | none | 0.2  | 91% | 100% | 100% | 99% |
| 30  | 0.8 | high  | none | 0.4  | 100%| 100% | 100% | 100% |
| 30  | 0.8 | high  | med  | 0.0  | 92% | 100% | 100% | 87% |
| 30  | 0.8 | high  | med  | 0.2  | 85% | 100% | 100% | 99% |
| 30  | 0.8 | high  | med  | 0.4  | 96% | 100% | 100% | 100% |
| 30  | 0.8 | high  | high | 0.0  | 92% | 100% | 100% | 88% |
| 30  | 0.8 | high  | high | 0.2  | 84% | 100% | 100% | 98% |
| 30  | 0.8 | high  | high | 0.4  | 95% | 100% | 100% | 100% |
| 60  | 0.0 | none  | none | 0.0  | 98% | 78%  | 78%  | 72% |
| 60 0.0 none none | 0.2 | 100% | 100% | 100% | 100% |
| 60 0.0 none none | 0.4 | 100% | 100% | 100% | 100% |
| 60 0.0 none med | 0.0 | 92% | 100% | 100% | 79% |
| 60 0.0 none med | 0.2 | 100% | 100% | 100% | 100% |
| 60 0.0 none high | 0.0 | 90% | 100% | 100% | 74% |
| 60 0.0 none high | 0.2 | 100% | 100% | 100% | 100% |
| 60 0.0 none high | 0.4 | 100% | 100% | 100% | 100% |
| 60 0.0 med none | 0.0 | 99% | 100% | 100% | 76% |
| 60 0.0 med none | 0.2 | 100% | 100% | 100% | 100% |
| 60 0.0 med med | 0.0 | 100% | 100% | 100% | 100% |
| 60 0.0 med med | 0.2 | 100% | 100% | 100% | 100% |
| 60 0.0 med high | 0.0 | 92% | 100% | 100% | 79% |
| 60 0.0 med high | 0.2 | 100% | 100% | 100% | 100% |
| 60 0.0 med high | 0.4 | 100% | 100% | 100% | 100% |
| 60 0.0 med high | 0.0 | 81% | 100% | 100% | 79% |
| 60 0.0 med high | 0.2 | 96% | 100% | 100% | 98% |
| 60 0.0 med high | 0.4 | 100% | 100% | 100% | 100% |
| 60 0.0 high none | 0.0 | 100% | 100% | 100% | 100% |
| 60 0.0 high none | 0.2 | 100% | 100% | 100% | 100% |
| 60 0.0 high none | 0.4 | 100% | 100% | 100% | 100% |
| 60 0.0 high med | 0.0 | 100% | 100% | 100% | 100% |
| 60 0.0 high med | 0.2 | 100% | 100% | 100% | 100% |
| 60 0.0 high med | 0.4 | 100% | 100% | 100% | 100% |
| 60 0.0 high high | 0.0 | 97% | 100% | 100% | 97% |
| 60 0.0 high high | 0.2 | 90% | 100% | 100% | 97% |
| 60 0.0 high high | 0.4 | 99% | 100% | 100% | 100% |
| 60 0.2 none none | 0.0 | 98% | 100% | 100% | 76% |
| 60 0.2 none none | 0.2 | 100% | 100% | 100% | 100% |
| 60 0.2 none none | 0.4 | 100% | 100% | 100% | 100% |
| 60 0.2 none med | 0.0 | 95% | 100% | 100% | 42% |
| 60 0.2 none med | 0.2 | 99% | 100% | 100% | 99% |
| 60 0.2 none med | 0.4 | 100% | 100% | 100% | 100% |
| 60 0.2 none med | 0.0 | 94% | 100% | 100% | 33% |
| 60 0.2 none med | 0.2 | 96% | 100% | 100% | 100% |
| 60 0.2 none med | 0.4 | 99% | 100% | 100% | 100% |
| 60 0.2 med none | 0.0 | 99% | 100% | 100% | 76% |
| 60 0.2 med none | 0.2 | 100% | 100% | 100% | 100% |
| 60 0.2 med none | 0.4 | 100% | 100% | 100% | 100% |
| 60 0.2 med med | 0.0 | 89% | 100% | 100% | 55% |
| 60 0.2 med med | 0.2 | 94% | 100% | 100% | 99% |
| 60 0.2 med med | 0.4 | 100% | 100% | 100% | 100% |
| 60 0.2 med med | 0.0 | 88% | 100% | 100% | 53% |
| 60 0.2 med med | 0.2 | 78% | 100% | 100% | 98% |
| 60 0.2 med med | 0.4 | 93% | 100% | 100% | 100% |
| 60 0.2 med none | 0.0 | 98% | 100% | 100% | 81% |
| 60 0.2 med none | 0.2 | 100% | 100% | 100% | 99% |
| 60 0.2 med none | 0.4 | 100% | 100% | 100% | 100% |
| 60 0.2 high med | 0.0 | 85% | 100% | 100% | 68% |
| 60 0.2 high med | 0.2 | 93% | 100% | 100% | 98% |
| 60 0.2 high med | 0.4 | 100% | 100% | 100% | 100% |
| 60 0.2 high high | 0.0 | 82% | 100% | 100% | 58% |
| 60 0.2 high high | 0.2 | 73% | 100% | 100% | 97% |
| 60 0.2 high high | 0.4 | 90% | 100% | 100% | 100% |
| 60 0.5 none none | 0.0 | 98% | 100% | 100% | 80% |
| 60 0.5 none none | 0.2 | 100% | 100% | 100% | 100% |
| 60 0.5 none none | 0.4 | 100% | 100% | 100% | 100% |
| 60 0.5 none med | 0.0 | 93% | 100% | 100% | 65% |
| 60 0.5 none med | 0.2 | 95% | 100% | 100% | 100% |
| 60 0.5 none med | 0.4 | 100% | 100% | 100% | 100% |
| 60 0.5 none high | 0.0 | 92% | 100% | 100% | 60% |
| 60 0.5 none high | 0.2 | 93% | 100% | 100% | 100% |
| 60 0.5 none high | 0.4 | 99% | 100% | 100% | 100% |
| 60 0.5 med none | 0.0 | 92% | 100% | 100% | 75% |
| 60 0.5 med none | 0.2 | 98% | 100% | 100% | 100% |
| 60 | 0.5 | med | none | 0.4 | 100% | 100% | 100% | 100% |
| 60 | 0.5 | med | med | 0.0 | 88% | 100% | 100% | 74% |
| 60 | 0.5 | med | med | 0.2 | 80% | 100% | 100% | 99% |
| 60 | 0.5 | med | med | 0.4 | 97% | 100% | 100% | 100% |
| 60 | 0.5 | high | none | 0.0 | 87% | 100% | 100% | 68% |
| 60 | 0.5 | high | med | 0.2 | 70% | 100% | 100% | 100% |
| 60 | 0.5 | high | high | 0.4 | 88% | 100% | 100% | 100% |
| 60 | 0.5 | high | none | 0.0 | 91% | 100% | 100% | 77% |
| 60 | 0.5 | high | none | 0.2 | 97% | 100% | 100% | 100% |
| 60 | 0.5 | high | none | 0.4 | 100% | 100% | 100% | 100% |
| 60 | 0.5 | high | med | 0.0 | 87% | 100% | 100% | 72% |
| 60 | 0.5 | high | med | 0.2 | 71% | 100% | 100% | 100% |
| 60 | 0.5 | high | med | 0.4 | 93% | 100% | 100% | 100% |
| 60 | 0.5 | high | high | 0.0 | 87% | 100% | 100% | 71% |
| 60 | 0.5 | high | high | 0.2 | 64% | 100% | 100% | 100% |
| 60 | 0.5 | high | high | 0.4 | 85% | 100% | 100% | 100% |
| 60 | 0.8 | none | none | 0.0 | 97% | 100% | 100% | 80% |
| 60 | 0.8 | none | none | 0.2 | 100% | 100% | 100% | 100% |
| 60 | 0.8 | none | none | 0.4 | 100% | 100% | 100% | 100% |
| 60 | 0.8 | none | med | 0.0 | 96% | 100% | 100% | 76% |
| 60 | 0.8 | none | med | 0.2 | 96% | 100% | 100% | 100% |
| 60 | 0.8 | none | med | 0.4 | 100% | 100% | 100% | 100% |
| 60 | 0.8 | none | high | 0.0 | 94% | 100% | 100% | 100% |
| 60 | 0.8 | none | high | 0.2 | 93% | 100% | 100% | 100% |
| 60 | 0.8 | none | high | 0.4 | 99% | 100% | 100% | 100% |
| 60 | 0.8 | med | none | 0.0 | 94% | 100% | 100% | 79% |
| 60 | 0.8 | med | none | 0.2 | 95% | 100% | 100% | 100% |
| 60 | 0.8 | med | none | 0.4 | 100% | 100% | 100% | 100% |
| 60 | 0.8 | med | med | 0.0 | 93% | 100% | 100% | 100% |
| 60 | 0.8 | med | med | 0.2 | 87% | 100% | 100% | 100% |
| 60 | 0.8 | med | med | 0.4 | 98% | 100% | 100% | 100% |
| 60 | 0.8 | med | high | 0.0 | 92% | 100% | 100% | 77% |
| 60 | 0.8 | med | high | 0.2 | 85% | 100% | 100% | 100% |
| 60 | 0.8 | med | high | 0.4 | 96% | 100% | 100% | 100% |
| 60 | 0.8 | high | none | 0.0 | 92% | 100% | 100% | 78% |
| 60 | 0.8 | high | none | 0.2 | 88% | 100% | 100% | 100% |
| 60 | 0.8 | high | none | 0.4 | 100% | 100% | 100% | 100% |
| 60 | 0.8 | high | med | 0.0 | 92% | 100% | 100% | 77% |
| 60 | 0.8 | high | med | 0.2 | 82% | 100% | 100% | 100% |
| 60 | 0.8 | high | med | 0.4 | 96% | 100% | 100% | 100% |
| 60 | 0.8 | high | high | 0.0 | 90% | 100% | 100% | 100% |
| 60 | 0.8 | high | high | 0.2 | 81% | 100% | 100% | 100% |
| 60 | 0.8 | high | high | 0.4 | 95% | 100% | 100% | 100% |
| 100 | 0.0 | none | none | 0.0 | 98% | 92% | 92% | 64% |
| 100 | 0.0 | none | none | 0.2 | 100% | 100% | 100% | 100% |
| 100 | 0.0 | none | none | 0.4 | 100% | 100% | 100% | 100% |
| 100 | 0.0 | none | med | 0.0 | 94% | 100% | 100% | 67% |
| 100 | 0.0 | none | med | 0.2 | 100% | 100% | 100% | 100% |
| 100 | 0.0 | none | med | 0.4 | 100% | 100% | 100% | 100% |
| 100 | 0.0 | none | high | 0.0 | 91% | 100% | 100% | 61% |
| 100 | 0.0 | none | high | 0.2 | 100% | 100% | 100% | 100% |
| 100 | 0.0 | none | high | 0.4 | 100% | 100% | 100% | 100% |
| 100 | 0.0 | med | none | 0.0 | 100% | 100% | 100% | 58% |
| 100 | 0.0 | med | none | 0.2 | 100% | 100% | 100% | 100% |
| 100 | 0.0 | med | none | 0.4 | 100% | 100% | 100% | 100% |
| 100 | 0.0 | med | med | 0.0 | 100% | 100% | 100% | 100% |
| 100 | 0.0 | med | med | 0.2 | 100% | 100% | 100% | 100% |
| 100 | 0.0 | med | med | 0.4 | 100% | 100% | 100% | 100% |
| Value | Label 1 | Value | Label 2 | Value 1 | Value 2 | Value 3 | Value 4 | Value 5 |
|-------|---------|-------|---------|--------|--------|--------|--------|--------|
| 100 0.0 | high | med | 0.0 | 100% | 100% | 100% | 64% |
| 100 0.0 | high | med | 0.2 | 100% | 100% | 100% | 100% |
| 100 0.0 | high | med | 0.4 | 100% | 100% | 100% | 100% |
| 100 0.0 | high | high | 0.0 | 73% | 100% | 100% | 70% |
| 100 0.0 | high | high | 0.2 | 96% | 100% | 100% | 99% |
| 100 0.2 | none | none | 0.0 | 97% | 100% | 100% | 69% |
| 100 0.2 | none | none | 0.2 | 100% | 100% | 100% | 100% |
| 100 0.2 | none | none | 0.4 | 100% | 100% | 100% | 100% |
| 100 0.2 | none | med | 0.0 | 97% | 100% | 100% | 70% |
| 100 0.2 | none | med | 0.2 | 96% | 100% | 100% | 100% |
| 100 0.2 | none | med | 0.4 | 100% | 100% | 100% | 100% |
| 100 0.2 | high | med | 0.0 | 97% | 100% | 100% | 19% |
| 100 0.2 | high | med | 0.2 | 97% | 100% | 100% | 100% |
| 100 0.2 | high | med | 0.4 | 100% | 100% | 100% | 100% |
| 100 0.2 | high | high | 0.0 | 98% | 100% | 100% | 67% |
| 100 0.2 | high | high | 0.2 | 100% | 100% | 100% | 100% |
| 100 0.2 | high | high | 0.4 | 100% | 100% | 100% | 100% |
| 100 0.2 | high | high | 0.6 | 98% | 100% | 100% | 99% |
| 100 0.2 | high | high | 0.8 | 94% | 100% | 100% | 100% |
| 100 0.5 | none | none | 0.0 | 98% | 100% | 100% | 67% |
| 100 0.5 | none | none | 0.2 | 100% | 100% | 100% | 100% |
| 100 0.5 | none | none | 0.4 | 100% | 100% | 100% | 100% |
| 100 0.5 | none | med | 0.0 | 95% | 100% | 100% | 51% |
| 100 0.5 | none | med | 0.2 | 98% | 100% | 100% | 100% |
| 100 0.5 | none | med | 0.4 | 100% | 100% | 100% | 100% |
| 100 0.5 | none | high | 0.0 | 94% | 100% | 100% | 48% |
| 100 0.5 | none | high | 0.2 | 95% | 100% | 100% | 100% |
| 100 0.5 | none | high | 0.4 | 100% | 100% | 100% | 100% |
| 100 0.5 | med | none | 0.0 | 93% | 100% | 100% | 64% |
| 100 0.5 | med | none | 0.2 | 99% | 100% | 100% | 100% |
| 100 0.5 | med | none | 0.4 | 100% | 100% | 100% | 100% |
| 100 0.5 | med | med | 0.0 | 88% | 100% | 100% | 59% |
| 100 0.5 | med | med | 0.2 | 78% | 100% | 100% | 100% |
| 100 0.5 | med | med | 0.4 | 98% | 100% | 100% | 100% |
| 100 0.5 | med | high | 0.0 | 88% | 100% | 100% | 59% |
| 100 0.5 | med | high | 0.2 | 67% | 100% | 100% | 100% |
| 100 0.5 | med | high | 0.4 | 91% | 100% | 100% | 100% |
| 100 0.5 | high | none | 0.0 | 91% | 100% | 100% | 66% |
| 100 0.5 | high | none | 0.2 | 99% | 100% | 100% | 100% |
| 100 0.5 | high | none | 0.4 | 100% | 100% | 100% | 100% |
| 100 0.5 | high | med | 0.0 | 86% | 100% | 100% | 60% |
| 100 0.5 | high | med | 0.2 | 68% | 100% | 100% | 100% |
| 100 0.5 | high | med | 0.4 | 97% | 100% | 100% | 100% |
| 100 0.5 | high | high | 0.0 | 87% | 100% | 100% | 59% |
| 100 0.5 | high | high | 0.2 | 57% | 100% | 100% | 100% |
| 100 0.5 | high | high | 0.4 | 85% | 100% | 100% | 100% |
| 100 0.8 | none | none | 0.0 | 98% | 100% | 100% | 70% |
| 100 0.8 | none | none | 0.2 | 100% | 100% | 100% | 100% |
| 100 0.8 | none | none | 0.4 | 100% | 100% | 100% | 100% |
| 100 0.8 | none | med | 0.0 | 95% | 100% | 100% | 63% |
| QRP  | PB   | 0.0 | 0.2 | 0.4 | 94% | 100% | 100% | 100% | 100% | 95% | 100% | 100% | 100% | 59% | 100% | 100% | 100% | 100% | 92% | 100% | 100% | 100% | 93% | 100% | 100% | 100% | 96% | 100% | 100% | 100% | 93% | 100% | 100% | 100% | 97% | 100% | 100% | 100% | 96% | 100% | 100% | 100% | 94% | 100% | 100% | 100% | 96% | 100% | 100% | 100% | 94% | 100% | 100% | 100% | 96% | 100% | 100% | 100% | 96% | 100% | 100% | 100% | 98% | 100% | 100% | 100% | 91% | 100% | 100% | 100% | 92% | 100% | 100% | 100% | 92% | 100% | 100% | 100% | 91% | 100% | 100% | 100% | 98% | 100% | 100% | 100% | 92% | 100% | 100% | 100% | 81% | 100% | 100% | 100% | 94% | 100% | 100% | 100% | 92% | 100% | 100% | 100% | 82% | 100% | 100% | 100% | 92% | 100% | 100% | 100% | 98% | 100% | 100% | 100% | 82% | 100% | 100% | 100% | 92% | 100% | 100% | 100% | 82% | 100% | 100% | 100% | 92% | 100% | 100% | 100% |

Table 1

*QRP = QRP Environment, PB = Publication bias*
Figure 2. $p$-curve’s performance in estimating the true mean of all studies which have been submitted to the $p$-curve analysis. Figure available at https://osf.io/rf3ys, under a CC-BY4.0 license.
Figure 3. Rejection rates when $\delta = 0$ or $\delta = 0.20$. 
Figure 4. Rejection rates for hypothesis testing when $\delta = 0$ or $\delta = 0.80$. 
Figure 5. Estimation when $\delta = 0$ or $\delta = 0.20$. 

(A) no publication bias

(B) medium publication bias

(C) strong publication bias

Estimated effect size

QRP Env. ○ none □ med △ high δ @ 0 ● 0.2
Figure 6. Estimation when $\delta = 0$ or $\delta = 0.8$. 
**Figure 7.** Mean error (ME) for all methods with no publication bias. Color coding is as follows: darkest = |ME| < .1; medium = .1 ≤ |ME| < .15; lightest = .15 ≤ |ME|
Figure 8. Mean error ($ME$) for all methods with medium publication bias. Color coding is as follows: darkest = $|ME| < .1$; medium = $.1 \leq |ME| < .15$; lightest = $.15 \leq |ME|$. 

| $k$ | appEnv | RE | TF | PP | PC | PU | SP | WA | RE | TF | PP | PC | PU | SP | WA |
|-----|-------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 100 2 | $0.05 \pm 0.01$ | $0.22 \pm 0.01$ | $0.26 \pm 0.01$ | $0.24 \pm 0.01$ | $0.26 \pm 0.01$ | $0.20 \pm 0.01$ | $0.24 \pm 0.01$ | $0.20 \pm 0.01$ | $0.12 \pm 0.01$ | $0.12 \pm 0.01$ | $0.12 \pm 0.01$ | $0.12 \pm 0.01$ | $0.12 \pm 0.01$ | $0.12 \pm 0.01$ | $0.12 \pm 0.01$ |
| 60 2 | $0.15 \pm 0.02$ | $0.18 \pm 0.02$ | $0.20 \pm 0.02$ | $0.22 \pm 0.02$ | $0.21 \pm 0.02$ | $0.20 \pm 0.02$ | $0.18 \pm 0.02$ | $0.15 \pm 0.02$ | $0.12 \pm 0.02$ | $0.12 \pm 0.02$ | $0.12 \pm 0.02$ | $0.12 \pm 0.02$ | $0.12 \pm 0.02$ | $0.12 \pm 0.02$ |
| 30 2 | $0.23 \pm 0.03$ | $0.27 \pm 0.03$ | $0.27 \pm 0.03$ | $0.25 \pm 0.03$ | $0.24 \pm 0.03$ | $0.24 \pm 0.03$ | $0.22 \pm 0.03$ | $0.20 \pm 0.03$ | $0.18 \pm 0.03$ | $0.18 \pm 0.03$ | $0.18 \pm 0.03$ | $0.18 \pm 0.03$ | $0.18 \pm 0.03$ | $0.18 \pm 0.03$ |
| 20 2 | $0.26 \pm 0.04$ | $0.29 \pm 0.04$ | $0.30 \pm 0.04$ | $0.29 \pm 0.04$ | $0.27 \pm 0.04$ | $0.27 \pm 0.04$ | $0.25 \pm 0.04$ | $0.23 \pm 0.04$ | $0.20 \pm 0.04$ | $0.19 \pm 0.04$ | $0.19 \pm 0.04$ | $0.19 \pm 0.04$ | $0.19 \pm 0.04$ | $0.19 \pm 0.04$ |
| 100 1 | $0.02 \pm 0.01$ | $0.03 \pm 0.01$ | $0.03 \pm 0.01$ | $0.04 \pm 0.01$ | $0.04 \pm 0.01$ | $0.04 \pm 0.01$ | $0.04 \pm 0.01$ | $0.04 \pm 0.01$ | $0.04 \pm 0.01$ | $0.04 \pm 0.01$ | $0.04 \pm 0.01$ | $0.04 \pm 0.01$ | $0.04 \pm 0.01$ | $0.04 \pm 0.01$ |
| 30 1 | $0.04 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.04 \pm 0.02$ | $0.04 \pm 0.02$ | $0.04 \pm 0.02$ | $0.04 \pm 0.02$ | $0.04 \pm 0.02$ | $0.04 \pm 0.02$ | $0.04 \pm 0.02$ | $0.04 \pm 0.02$ | $0.04 \pm 0.02$ | $0.04 \pm 0.02$ | $0.04 \pm 0.02$ |
| 10 1 | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ |
| 100 0 | $0.02 \pm 0.01$ | $0.02 \pm 0.01$ | $0.02 \pm 0.01$ | $0.02 \pm 0.01$ | $0.02 \pm 0.01$ | $0.02 \pm 0.01$ | $0.02 \pm 0.01$ | $0.02 \pm 0.01$ | $0.02 \pm 0.01$ | $0.02 \pm 0.01$ | $0.02 \pm 0.01$ | $0.02 \pm 0.01$ | $0.02 \pm 0.01$ | $0.02 \pm 0.01$ |
| 60 0 | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ |
| 30 0 | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ |
| 10 0 | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ | $0.05 \pm 0.02$ |

$k$ appEnv: RE TF PP PC PU SP WA RE TF PP PC PU SP WA RE TF PP PC PU SP WA RE TF PP PC PU SP WA RE TF PP PC PU SP WA
Figure 9. Mean error (ME) for all methods with strong publication bias. Color coding is as follows: darkest = |ME| < 1; medium = 1 ≤ |ME| < 1.5; lightest = 1.5 ≤ |ME|.
**Figure 10.** Root mean squared error (RMSE) for all methods with no publication bias. Color coding is as follows: darkest = \( RMSE < .1 \); medium = \( .1 \leq RMSE < .15 \); lightest = \( .15 \leq RMSE \).

| Method | \( \tau = 0.0 \) | \( \tau = 0.2 \) | \( \tau = 0.4 \) |
|--------|----------------|----------------|----------------|
| 100 2  | 0.03 ± 0.02 | 0.04 ± 0.04 | 0.05 ± 0.04 |
| 60 2   | 0.03 ± 0.03 | 0.05 ± 0.05 | 0.06 ± 0.05 |
| 30 2   | 0.04 ± 0.04 | 0.06 ± 0.06 | 0.07 ± 0.06 |
| 10 2   | 0.05 ± 0.07 | 0.17 ± 0.13 | 0.13 ± 0.13 |
| 100 1  | 0.05 ± 0.05 | 0.04 ± 0.04 | 0.03 ± 0.04 |
| 60 1   | 0.06 ± 0.06 | 0.12 ± 0.12 | 0.14 ± 0.14 |
| 30 1   | 0.07 ± 0.07 | 0.19 ± 0.13 | 0.11 ± 0.11 |
| 10 1   | 0.08 ± 0.08 | 0.32 ± 0.14 | 0.19 ± 0.19 |
| 100 0  | 0.09 ± 0.09 | 0.09 ± 0.09 | 0.09 ± 0.09 |
| 60 0   | 0.10 ± 0.10 | 0.20 ± 0.10 | 0.11 ± 0.11 |
| 30 0   | 0.11 ± 0.11 | 0.30 ± 0.11 | 0.12 ± 0.12 |
| 10 0   | 0.12 ± 0.12 | 0.15 ± 0.15 | 0.15 ± 0.15 |

\( \delta = 0.5 \)

\( \delta = 0.2 \)

\( \delta = 0.0 \)
\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{supplemental_material.png}
\caption{Root mean squared error (RMSE) for all methods with medium publication bias. Color coding is as follows: darkest = RMSE < .1; medium = .1 \leq RMSE < .15; lightest = .15 \leq RMSE.}
\end{figure}
**Figure 12.** Root mean squared error (RMSE) for all methods with strong publication bias.
Color coding is as follows: darkest = RMSE < .1; medium = .1 ≤ RMSE < .15; lightest = .15 ≤ RMSE.

Supplemental Material 27
| Sample Size | 0.82 | 0.9 | 0.61 | 0.9 | 0.61 | 0.85 | 0.93 | 0.61 | 0.55 | 0.73 | 0.69 | 0.78 | 0.8 | 0.25 | 0.48 | 0.9 | 0.2 | 0.68 |
|-------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 100         | 0.82 | 0.9 | 0.61 | 0.9 | 0.61 | 0.85 | 0.93 | 0.61 | 0.55 | 0.73 | 0.69 | 0.78 | 0.8 | 0.25 | 0.48 | 0.9 | 0.2 | 0.68 |
| 60          | 0.87 | 0.89 | 0.69 | 0.74 | 0.87 | 0.93 | 0.81 | 0.75 | 0.58 | 0.78 | 0.84 | 0.78 | 0.89 | 0.47 | 0.49 | 0.86 | 0.51 | 0.72 |
| 30          | 0.82 | 0.83 | 0.81 | 0.66 | 0.9 | 0.94 | 0.82 | 0.89 | 0.78 | 0.86 | 0.88 | 0.84 | 0.9 | 0.73 | 0.52 | 0.86 | 0.82 | 0.82 |
| 10          | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 |
| 100         | 0.91 | 0.73 | 0.76 | 0.89 | 0.93 | 0.84 | 0.68 | 0.58 | 0.64 | 0.77 | 0.77 | 0.86 | 0.31 | 0.52 | 0.83 | 0.4 | 0.72 |
| 60          | 0.93 | 0.94 | 0.87 | 0.9 | 0.94 | 0.94 | 0.91 | 0.82 | 0.66 | 0.89 | 0.91 | 0.82 | 0.93 | 0.75 | 0.6 | 0.89 | 0.86 | 0.82 |
| 30          | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 |
| 10          | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 |

$\tau = 0.0$

$\tau = 0.2$

$\tau = 0.4$

$\delta = 0.8$

$\delta = 0.5$

$\delta = 0.2$

$\delta = 0.0$

---

Figure 13. Coverage probability ($CP$) for all methods with no publication bias. Color coding is as follows: darkest = $|CP - 0.95| < 0.01$; medium = $0.01 \leq |CP - 0.95| < 0.02$; lightest = $0.02 \leq |CP - 0.95|$. 

SUPPLEMENTAL MATERIAL 28
Figure 14. Coverage probability (CP) for all methods with medium publication bias. Color coding is as follows: darkest: $|CP-0.95| < .01$; medium: $0.01 \leq |CP-0.95| < .02$; lightest: $0.02 \leq |CP-0.95|$. 

$\tau = 0.0$  
$\tau = 0.2$  
$\tau = 0.4$  
$
\begin{array}{ccc|ccc|ccc}
\tau & 0.0 & 0.2 & 0.4 & 0.0 & 0.2 & 0.4 & 0.0 & 0.2 & 0.4 \\
100 & 0.81 & 0.85 & 0.83 & 0.81 & 0.83 & 0.86 & 0.83 & 0.86 & 0.86 \\
70 & 0.77 & 0.81 & 0.80 & 0.77 & 0.81 & 0.84 & 0.81 & 0.84 & 0.84 \\
50 & 0.73 & 0.77 & 0.76 & 0.73 & 0.76 & 0.80 & 0.76 & 0.80 & 0.80 \\
30 & 0.69 & 0.73 & 0.72 & 0.69 & 0.72 & 0.77 & 0.72 & 0.77 & 0.77 \\
10 & 0.65 & 0.69 & 0.69 & 0.65 & 0.69 & 0.74 & 0.69 & 0.74 & 0.74 \\
\end{array}$
Supplemental Material

| k | q, p, Env | RE | TF | PP | PC | 3P | WA | RE | TF | PP | PC | 3P | WA | RE | TF | PP | PC | 3P | WA | RE | TF | PP | PC | 3P | WA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 0.86 | 0.93 | 0.92 | 0.93 | 0.93 | 0.91 | 0.87 | 0.84 | 0.9 | 0.94 | 0.91 | 0.85 | 0.65 | 0.83 | 0.86 | 0.98 | 0.88 | 0.91 | 0.86 | 0.95 | 0.83 | 0.92 | 0.76 |

Figure 15. Coverage probability (CP) for all methods with strong publication bias. Color coding is as follows: darkest = |CP - 0.95| < 0.01; medium = |CP - 0.95| < 0.02; lightest = 0.02 ≤ |CP - 0.95| .
Supplemental Material

Table 16. Null hypothesis rejection rates ($H_0RR$) for all methods with no publication bias. Color coding is as follows: darkest = $H_0RR < .50$; medium = $.50 \leq H_0RR < .80$; lightest = $.80 \leq H_0RR$. Note: When this $\delta > 0$, $H_0RR$ is statistical power; when $\delta = 0$, $H_0RR$ is Type I error or the false positive rate.

| $k$ | $\tau = 0.0$ | $\tau = 0.2$ | $\tau = 0.4$ |
|-----|--------------|--------------|--------------|
|      | RE TF PP PC 3P WA | RE TF PP PC 3P WA | RE TF PP PC 3P WA |
| 100 | 2 | 1 | 1 | 0.94 | 1 |
| 60  | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 0.97 | 0.82 | 0.99 |
| 30  | 1 | 1 | 1 | 0.97 | 1 | 1 | 1 | 0.97 | 0.82 | 0.99 |
| 10  | 1 | 1 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| 100 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.97 | 0.82 | 0.99 |
| 60  | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.97 | 0.82 | 0.99 |
| 30  | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.97 | 0.82 | 0.99 |
| 10  | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.97 | 0.82 | 0.99 |
| 100 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0.97 | 0.82 | 0.99 |
| 60  | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0.97 | 0.82 | 0.99 |
| 30  | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0.97 | 0.82 | 0.99 |
| 10  | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0.97 | 0.82 | 0.99 |
| $k$ | $\text{appEnv}$ | RE TF PP PC 3P WA | RE TF PP PC 3P WA | RE TF PP PC 3P WA |
| 100 | 2 | 1 | 1 | 0.94 | 1 |
| 60  | 2 | 1 | 1 | 1 | 1 | 0.94 | 1 |
| 30  | 2 | 1 | 1 | 0.94 | 1 | 1 | 1 | 0.94 | 1 |
| 10  | 2 | 1 | 1 | 0.94 | 1 | 1 | 1 | 0.94 | 1 |
| 100 | 1 | 1 | 1 | 1 | 1 | 0.94 | 1 | 1 | 1 | 0.94 | 1 |
| 60  | 1 | 1 | 1 | 1 | 1 | 0.94 | 1 | 1 | 1 | 0.94 | 1 |
| 30  | 1 | 1 | 1 | 1 | 1 | 0.94 | 1 | 1 | 1 | 0.94 | 1 |
| 10  | 1 | 1 | 1 | 1 | 1 | 0.94 | 1 | 1 | 1 | 0.94 | 1 |

Note: When this $\delta > 0$, $H_0RR$ is statistical power; when $\delta = 0$, $H_0RR$ is Type I error or the false positive rate.
| $\tau$ | $\delta = 0.0$ | $\delta = 0.2$ | $\delta = 0.5$ |
|-------|----------------|----------------|---------------|
| 100 2 | 1 1 1 1 1 1 | 1 1 1 1 1 1 | 1 1 1 1 1 1 |
| 60 2  | 1 1 1 1 1 1 | 1 1 1 1 1 1 | 1 1 1 1 1 1 |
| 30 2  | 1 1 1 1 1 1 | 1 1 1 1 0.99 | 1 1 1 1 0.97 |
| 10 2  | 1 1 0.99 0.99 1 | 1 1 1 1 0.98 | 1 1 1 1 0.94 |
| 100 1 | 1 1 1 1 1 1 | 1 1 1 1 1 1 | 1 1 1 1 1 1 |
| 60 1  | 1 1 1 1 1 1 | 1 1 1 1 1 1 | 1 1 1 1 1 1 |
| 30 1  | 1 1 1 0.99 0.99 | 1 1 1 1 0.98 | 1 1 1 1 0.94 |
| 10 1  | 1 1 1 1 1 1 | 1 1 1 1 1 1 | 1 1 1 1 1 1 |
| 30 0  | 1 1 1 1 1 1 | 1 1 1 1 1 1 | 1 1 1 1 1 1 |
| 10 0  | 1 1 1 1 1 1 | 1 1 1 1 1 1 | 1 1 1 1 1 1 |

$k$ qrpEnv | RE | TF | PP | PC | PU | 3P | WA | RE | TF | PP | PC | PU | 3P | WA | RE | TF | PP | PC | PU | 3P | WA
|-------|----------------|----------------|---------------|-------|----------------|----------------|---------------|-------|----------------|----------------|---------------|-------|----------------|----------------|---------------|-------|----------------|----------------|---------------|-------|----------------|----------------|---------------|
| 100 2 | 1 1 0.96 0.66 | 1 1 0.89 0.27 | 1 1 0.96 0.27 | 1 1 0.96 0.27 |
| 60 2 | 1 1 0.88 0.51 | 1 1 0.82 0.24 | 1 1 0.87 0.24 | 1 1 0.87 0.24 |
| 30 2 | 1 1 0.99 0.67 | 1 1 0.99 0.27 | 1 1 0.99 0.27 | 1 1 0.99 0.27 |
| 10 2 | 1 1 0.99 0.67 | 1 1 0.99 0.27 | 1 1 0.99 0.27 | 1 1 0.99 0.27 |
| 100 1 | 1 1 0.99 0.67 | 1 1 0.99 0.27 | 1 1 0.99 0.27 | 1 1 0.99 0.27 |
| 60 1 | 1 1 0.99 0.67 | 1 1 0.99 0.27 | 1 1 0.99 0.27 | 1 1 0.99 0.27 |
| 30 1 | 1 1 0.99 0.67 | 1 1 0.99 0.27 | 1 1 0.99 0.27 | 1 1 0.99 0.27 |
| 10 1 | 1 1 0.99 0.67 | 1 1 0.99 0.27 | 1 1 0.99 0.27 | 1 1 0.99 0.27 |
| 30 0 | 1 1 0.99 0.67 | 1 1 0.99 0.27 | 1 1 0.99 0.27 | 1 1 0.99 0.27 |
| 10 0 | 1 1 0.99 0.67 | 1 1 0.99 0.27 | 1 1 0.99 0.27 | 1 1 0.99 0.27 |

$k$ qrpEnv | RE | TF | PP | PC | PU | 3P | WA | RE | TF | PP | PC | PU | 3P | WA | RE | TF | PP | PC | PU | 3P | WA
|-------|----------------|----------------|---------------|-------|----------------|----------------|---------------|-------|----------------|----------------|---------------|-------|----------------|----------------|---------------|-------|----------------|----------------|---------------|-------|----------------|----------------|---------------|
| 100 2 | 1 1 0.60 0.77 | 1 1 0.60 0.77 | 1 1 0.60 0.77 | 1 1 0.60 0.77 |
| 60 2 | 1 1 0.60 0.77 | 1 1 0.60 0.77 | 1 1 0.60 0.77 | 1 1 0.60 0.77 |
| 30 2 | 1 1 0.60 0.77 | 1 1 0.60 0.77 | 1 1 0.60 0.77 | 1 1 0.60 0.77 |
| 10 2 | 1 1 0.60 0.77 | 1 1 0.60 0.77 | 1 1 0.60 0.77 | 1 1 0.60 0.77 |
| 100 1 | 1 1 0.60 0.77 | 1 1 0.60 0.77 | 1 1 0.60 0.77 | 1 1 0.60 0.77 |
| 60 1 | 1 1 0.60 0.77 | 1 1 0.60 0.77 | 1 1 0.60 0.77 | 1 1 0.60 0.77 |
| 30 1 | 1 1 0.60 0.77 | 1 1 0.60 0.77 | 1 1 0.60 0.77 | 1 1 0.60 0.77 |
| 10 1 | 1 1 0.60 0.77 | 1 1 0.60 0.77 | 1 1 0.60 0.77 | 1 1 0.60 0.77 |
| 30 0 | 1 1 0.60 0.77 | 1 1 0.60 0.77 | 1 1 0.60 0.77 | 1 1 0.60 0.77 |
| 10 0 | 1 1 0.60 0.77 | 1 1 0.60 0.77 | 1 1 0.60 0.77 | 1 1 0.60 0.77 |

Figure 17. Null hypothesis rejection rates (H0RR) for all methods with medium publication bias. Color coding is as follows: darkest = H0RR < .50; medium = .50 ≤ H0RR < .80; lightest = .80 ≤ H0RR. Note: When this $\delta > 0$, H0RR is statistical power; when $\delta = 0$, H0RR is Type I error or the false positive rate.
### Figure 18. Null hypothesis rejection rates (H₀RR) for all methods with strong publication bias. Color coding is as follows: darkest = H₀RR < .50; medium = .50 ≤ H₀RR < .80; lightest = .80 ≤ H₀RR. Note: When this δ > 0, H₀RR is statistical power; when δ = 0, H₀RR is Type I error or the false positive rate.

| k | Δ = 0.0 | τ = 0.2 | τ = 0.4 |
|---|---|---|---|
| 1 | 1.0 | 0.99 | 0.86 |
| 2 | 1.0 | 0.99 | 0.86 |
| 3 | 1.0 | 0.99 | 0.86 |
| 4 | 1.0 | 0.99 | 0.86 |
| 5 | 1.0 | 0.99 | 0.86 |
| 6 | 1.0 | 0.99 | 0.86 |
| 7 | 1.0 | 0.99 | 0.86 |
| 8 | 1.0 | 0.99 | 0.86 |
| 9 | 1.0 | 0.99 | 0.86 |
| 10 | 1.0 | 0.99 | 0.86 |

### Table 10: Null hypothesis rejection rates (H₀RR) for all methods with strong publication bias. Color coding is as follows: darkest = H₀RR < .50; medium = .50 ≤ H₀RR < .80; lightest = .80 ≤ H₀RR. Note: When this δ > 0, H₀RR is statistical power; when δ = 0, H₀RR is Type I error or the false positive rate.

| δ | 0.0 | 0.5 | 2.0 |
|---|---|---|---|
| 1 | 1.0 | 0.99 | 0.86 |
| 2 | 1.0 | 0.99 | 0.86 |
| 3 | 1.0 | 0.99 | 0.86 |
| 4 | 1.0 | 0.99 | 0.86 |
| 5 | 1.0 | 0.99 | 0.86 |
| 6 | 1.0 | 0.99 | 0.86 |
| 7 | 1.0 | 0.99 | 0.86 |
| 8 | 1.0 | 0.99 | 0.86 |
| 9 | 1.0 | 0.99 | 0.86 |
| 10 | 1.0 | 0.99 | 0.86 |

### Table 11: Null hypothesis rejection rates (H₀RR) for all methods with strong publication bias. Color coding is as follows: darkest = H₀RR < .50; medium = .50 ≤ H₀RR < .80; lightest = .80 ≤ H₀RR. Note: When this δ > 0, H₀RR is statistical power; when δ = 0, H₀RR is Type I error or the false positive rate.

| δ | 0.0 | 0.5 | 2.0 |
|---|---|---|---|
| 1 | 1.0 | 0.99 | 0.86 |
| 2 | 1.0 | 0.99 | 0.86 |
| 3 | 1.0 | 0.99 | 0.86 |
| 4 | 1.0 | 0.99 | 0.86 |
| 5 | 1.0 | 0.99 | 0.86 |
| 6 | 1.0 | 0.99 | 0.86 |
| 7 | 1.0 | 0.99 | 0.86 |
| 8 | 1.0 | 0.99 | 0.86 |
| 9 | 1.0 | 0.99 | 0.86 |
| 10 | 1.0 | 0.99 | 0.86 |