A Correction to Begg's Test for Publication Bias

Haben Michael[1] and Musie Ghebremichael[2]

[1]University of Massachusetts, Amherst, MA (hmichael@math.umass.edu)

[2]Ragon Institute and Harvard University, Cambridge, MA

(musie_ghebremichael@dfci.harvard.edu)

SUMMARY: Begg and Mazumdar proposed using a rank correlation test to test for publication bias when carrying out meta-analyses. The asymptotic variance of the rank correlation test statistic was derived under assumptions unmet by this application, often resulting in a loss of power. Low power when Begg's test is used to screen for publication bias may lead to false positives in a subsequent meta-analysis. We obtain the asymptotic bias under the common conditionally normal model as a function of the distribution of primary study variances. In simulations we consider the performance of Begg's test using an approximation to the correct asymptotic variance. We consider this performance under the common fixed effects and random effects frameworks. We then examine several meta-analyses drawn from the literature where the standard and bias-corrected versions of Begg's test lead to different conclusions.

KEYWORDS: Publication bias; Meta-analysis; Error rates; Kendall's tau.

1 Introduction

Meta-analysis is a popular technique for summarizing a body of studies. Key to the soundness of the approach is that the body of studies used in forming the summary be representative of the studies conducted. This requirement may fail to be met when publication bias is present, that is, when the availability of a study is tied to its findings (Begg 1994). Several hypothesis tests have been proposed with the goal of identifying the presence of publication bias on the basis of the relationship between the conclusion of the study and various study characteristics.

Common to these tests is that the null is held to be no publication bias. A typical conservative analyst might be expected to treat the presence of publication bias as the null. Given the manifold sources of publication bias, devising a test under such a null does not appear practical. The result of taking the null to be the absence of publication bias, however, is that Type II errors in the test for publication bias will often correspond to Type I errors in the subsequent meta-analysis. Assessing and improving the power of the screening test is therefore worthwhile.

One of the most common tests for publication bias, Begg's test (Begg and Mazumdar 1994), tests for correlation between the studies' reported effect sizes and their standard errors. An issue with Begg's test procedure is that it uses the asymptotic variance for a general correlation test derived under assumptions unmet by Begg's test. This nominal variance is often larger than the correct variance, as discussed below. As a result, the test does not reject as frequently as it ought, which, as mentioned above, is likely to lead to Type I error in the meta-analysis for which the publication bias test is being performed.

This issue with Begg's test has been noted previously, including by the author of the test (Begg 1994; Begg and Mazumdar 1994). More recently, Gjerdevik and Heuch (2014) showed that the observations forming the input to the rank correlation test are correlated, so that the usual assumptions for the test are not met. Since the rank correlation test depends on an asymptotic approximation, this criticism isn't entirely fair unless the effect of the correlation does not vanish in the limit, as we show.

The remainder of the paper is organized as follows. In Sections 2.1 and 2.2, we describe Begg's test in greater detail and identify the source of bias. In 2.3 we discuss the direction of the bias and relate it to the error rates of the test, and in Section 2.4 we present the form of the bias in the normal model assumed by Begg and Mazumdar (1994). In Section 3, we contrast the power of Begg's test as standardly used and a debiased version, using simulated data. In Section 4, we examine three meta-analysis drawn from biomedical literature in which the standard and debiased versions of Begg's test, applied in a hypothesis testing framework, offer diverging conclusions.

2 Asymptotic bias of Begg's test

2.1 Description of test

Begg's test is a test of correlation between the reported effect sizes and their reported variances. The premise is that a tendency to publish larger effect sizes induces a trend in effect sizes across their variances, and no such trend exists without selection. See Fig. 1 for an illustration.

The data consists of independent pairs $(Y_1, \sigma_1), \ldots, (Y_n, \sigma_n)$ representing the estimated effect sizes and sampling variances of n studies with a common mean effect size, say θ:

$$\begin{aligned} \mathrm{E}(Y_j \mid \sigma_j) &= \theta, j = 1, \ldots, n \\ \mathrm{Var}(Y_j \mid \sigma_j^2) &= \sigma_j^2. \end{aligned} \tag{1}$$

The null is that Y_j is uncorrelated with σ_j, $j = 1, \ldots, n$. The test statistic is Kendall's rank correlation coefficient,

$$\tau = 2\binom{n}{2}^{-1} \sum_{j<k} \{(U_j - U_k)(V_j - V_k) > 0\} - 1,$$

applied to the sequence of pairs (U_j, V_j) given by the data after standardizing the effect sizes,

$$\begin{aligned}(U_j, V_j) &= \left(\frac{Y_j - \hat{\theta}}{\sqrt{\sigma_j^2 - \sigma_{\hat{\theta}}^2}}, \sigma_j\right), j = 1, \ldots, n,\\ &\text{where}\\ \hat{\theta} &= (\sum_{j=1}^{n} Y_j/\sigma_j^2)/(\sum_{j=1}^{n} 1/\sigma_j^2),\\ \sigma_{\hat{\theta}}^2 &= 1/\sum_{j=1}^{n}(1/\sigma_j^2).\end{aligned} \tag{2}$$

The mean estimate $\hat{\theta} = (\sum_{j=1}^{n} Y_j/\sigma_j^2)/(\sum_{j=1}^{n} 1/\sigma_j^2)$ is the inverse-variance weighted estimate of the common study mean θ and $\sigma_{\hat{\theta}}^2 = 1/\sum_{j=1}^{n}(1/\sigma_j^2)$ is its variance, both conditional on the study variances. The test statistic counts the number of corresponding pairs of studentized effect sizes $U_j = (Y_j - \hat{\theta})/\sqrt{\sigma_j^2 - \sigma_{\hat{\theta}}^2}$ and variances $V_j = \sigma_j$ that concord in the sense that either $U_j < U_k$ and $V_j < V_k$ or $U_j > U_k$ and $V_j > V_k$. The null of no correlation is to be interpreted as no publication bias, and is rejected at level α when $\sqrt{9n/4}|\tau| > \Phi^{-1}(1-\alpha/2)$.

The asymptotic null variance 4/9 is derived under the assumption that the pairs form an IID sequence. This assumption does not hold for the pairs (2) due to the common terms $\hat{\theta}$ and $\sigma_{\hat{\theta}}^2$. While the latter is of order $1/n$ and typically negligible in the limit, the dependence induced by the summary statistic $\hat{\theta}$, ordinarily of order $1/\sqrt{n}$, must be accounted for in computing the asymptotic null variance of $\sqrt{n}\tau$.

2.2 Source of bias

The variance of $\sqrt{n}\tau$ is

$$\text{Var}(\sqrt{n}\tau) = 4n\binom{n}{2}^{-2} \sum_{i<j,k<l} \text{Cov}(\{(U_i - U_j)(V_i - V_j) > 0\}, \{(U_k - U_l)(V_k - V_l) > 0\}).$$

The sum has $\binom{n}{2}\binom{n-2}{2}$ terms where i, j, k, l are all distinct, $2(n-2)\binom{n}{2}$ terms where the set $\{i, j, k, l\}$ has size 3, and $\binom{n}{2}$ terms where $|\{i, j, k, l\}| = 2$, so

$$\begin{aligned}\mathrm{Var}(\sqrt{n}\tau) = 4\frac{(n-2)(n-3)}{n-1} \mathrm{Cov}(\{(U_1 - U_2)(V_1 - V_2) > 0\}, \{(U_3 - U_4)(V_3 - V_4) > 0\}) + \\ 16\frac{n-2}{n-1} \mathrm{Cov}(\{(U_1 - U_2)(V_1 - V_2) > 0\}, \{(U_1 - U_3)(V_1 - V_3) > 0\}) + O(1/n).\end{aligned}$$

The second term on the right-hand side, with the $O(1)$ coefficient, converges in probability to

$$\begin{aligned}&16 \cdot \mathrm{Cov}\left(\left\{\left(\frac{Y_1-\theta}{\sigma_1} - \frac{Y_2-\theta}{\sigma_2}\right)(\sigma_1 - \sigma_2) > 0\right\}, \left\{\left(\frac{Y_1-\theta}{\sigma_1} - \frac{Y_3-\theta}{\sigma_3}\right)(\sigma_1 - \sigma_3) > 0\right\}\right) \\ &\quad = 4/9,\end{aligned}$$

the usual asymptotic variance of Kendall's τ under the null that U_j and V_j are independent, $j = 1, \ldots, n$. The first term on the right-hand side, with an $O(n)$ coefficient, is a source of bias if the covariance

$$\begin{aligned}&\mathrm{Cov}(\{(U_1 - U_2)(V_1 - V_2) > 0\}, \{(U_3 - U_4)(V_3 - V_4) > 0\}) = \\ &\mathrm{Cov}\left(\left\{\left(\frac{Y_1 - \hat{\theta}}{\sqrt{\sigma_1^2 - \sigma_{\hat{\theta}}^2}} - \frac{Y_2 - \hat{\theta}}{\sqrt{\sigma_2^2 - \sigma_{\hat{\theta}}^2}}\right)(\sigma_1 - \sigma_2) > 0\right\}, \left\{\left(\frac{Y_3 - \hat{\theta}}{\sqrt{\sigma_3^2 - \sigma_{\hat{\theta}}^2}} - \frac{Y_4 - \hat{\theta}}{\sqrt{\sigma_4^2 - \sigma_{\hat{\theta}}^2}}\right)(\sigma_3 - \sigma_4) > 0\right\}\right)\end{aligned} \tag{3}$$

does not vanish faster than $1/n$. The false positive rate of Begg's test will exceed or fall below the nominal level when the direction of the bias is negative or positive, respectively, i.e., when the covariance (3) is positive or negative.

2.3 Direction of bias

Shifting the data if necessary, assume in (1) that the common mean θ is 0, and suppose that Y_j/σ_j has a fixed distribution, say F_Z, i.e., $Y_1, \ldots, Y_n$ belong to a scale family. Assume further that F_Z is the distribution of a symmetric random variable. Let $S^2 = 1/\sigma^2$ denote the study precisions. In summary,

$$\begin{aligned}
Z_1, \ldots, Z_n &\overset{IID}{\sim} F_Z \\
S_1, \ldots, S_n &\overset{IID}{\sim} F_S \\
Z_j &\sim -Z_j, \\
Z_j \mid S_j &\sim Z_j, \\
Y_j &= Z_j/S_j, j = 1, \ldots, n.
\end{aligned}$$

Ignoring the $O(1/n)$ terms $\sigma^2_{\hat{\theta}}$ in (2), the test statistic may be written in terms of Z and S as

$$\tau = \sum_{j<k} \left\{ \frac{Z_j - Z_k}{S_j - S_k} > \hat{\theta} \right\}.$$

The covariance (3) determining the bias of the asymptotic variance relative to 9/4 is

$$\mathrm{Cov}\left(\left\{ \frac{Z_1 - Z_2}{S1 - S2} > \hat{\theta} \right\}, \left\{ \frac{Z_3 - Z_4}{S3 - S4} > \hat{\theta} \right\} \right).$$

The grand mean estimate $\hat{\theta} = \sum_j Z_j S_j / \sum S_j^2$ induces dependence between the two terms in the covariance. By the symmetry of F_Z, $P(\{\frac{z_j - z_k}{s_j - s_k} > \hat{\theta}\} = 1/2$, so the last expression is

$$\begin{aligned}
&\mathrm{E}\left(\left\{ \frac{Z_1 - Z_2}{S1 - S2} > \hat{\theta} \right\} \left\{ \frac{Z_3 - Z_4}{S3 - S4} > \hat{\theta} \right\} \right) - 1/4 \\
&= P\left(\frac{Z_1 - Z_2}{S1 - S2} \wedge \frac{Z_3 - Z_4}{S3 - S4} > \hat{\theta} \right) - 1/4.
\end{aligned}$$

The symmetry of F_Z further implies that $\{\frac{Z_1 - Z_2}{S1 - S2} > \hat{\theta}\}\{\frac{Z_3 - Z_4}{S3 - S4} > \hat{\theta}\}$ has the same distribution as $\{\frac{Z_1 - Z_2}{S1 - S2} < \hat{\theta}\}\{\frac{Z_3 - Z_4}{S3 - S4} < \hat{\theta}\}$, so the condition for a positive bias, $\mathrm{E}(\{\frac{Z_1 - Z_2}{S1 - S2} > \hat{\theta}\}\{\frac{Z_3 - Z_4}{S3 - S4} >$

$\hat{\theta}\}) < 1/4$, is

$$\begin{aligned}
1/2 &> P\left(\left\{\frac{Z_1 - Z_2}{S1 - S2} > \hat{\theta}\right\}\left\{\frac{Z_3 - Z_4}{S3 - S4} > \hat{\theta}\right\}\right) + P\left(\left\{\frac{Z_1 - Z_2}{S1 - S2} < \hat{\theta}\right\}\left\{\frac{Z_3 - Z_4}{S3 - S4} < \hat{\theta}\right\}\right) \\
&= P(\mathbf{Z}_n^T(\mathbf{R}_1\mathbf{R}_2^T + \mathbf{R}_2\mathbf{R}_1^T)\mathbf{Z}_n/2 > 0),
\end{aligned}$$

where $\mathbf{Z}_n = (Z_1, \ldots, Z_n)$ and (for $n \geq 5$)

$$\begin{aligned}
\mathbf{R}_1 &= \left(\frac{\sum S_j^2}{S_1 - S_2} - S_1, \frac{-\sum S_j^2}{S_1 - S_2} - S_2, -S_3, \ldots, -S_n\right) / \sum S_j^2 \\
\mathbf{R}_2 &= \left(-S_1, -S_2, \frac{\sum S_j^2}{S_3 - S_4} - S_3, \frac{-\sum S_j^2}{S_3 - S_4} - S_4, -S_5, \ldots, -S_n\right) / \sum S_j^2.
\end{aligned}$$

The two nonzero eigenvalues of $(\mathbf{R}_1\mathbf{R}_2^T + \mathbf{R}_2\mathbf{R}_1^T)/2$ are $\lambda_a \pm \lambda_b$ where

$$\begin{aligned}
\lambda_a &= -1/(2\sum_j S_j^2) \\
\lambda_b &= \frac{\sqrt{((S_1 - S_2)^2/\sum_j S_j^2 - 2)((S_3 - S_4)^2/\sum_j S_j^2 - 2)}}{2(S_1 - S_2)(S_3 - S_4)}.
\end{aligned}$$

Then $|\lambda_a| < |\lambda_b|$, $\lambda_a - |\lambda_b| < 0 < \lambda_a + |\lambda_b|$, so one eigenvalue is negative and the other positive, while $|\lambda_b| - \lambda_a > |\lambda_b| + \lambda_a$, so the negative eigenvalue is larger in magnitude. Let $\mathbf{v}_1, \mathbf{v}_2$ denote unit eigenvectors associated respectively to the positive and negative eigenvalues. The condition for a positive bias takes the form

$$P\left(-\frac{\lambda_a}{\lambda_b} > \frac{\mathbf{Z}_n^T(\mathbf{v}_1^{\otimes 2} - \mathbf{v}_2^{\otimes 2})\mathbf{Z}_n}{\mathbf{Z}_n^T(\mathbf{v}_1^{\otimes 2} + \mathbf{v}_2^{\otimes 2})\mathbf{Z}_n}\right) > 1/2. \tag{4}$$

The ratio $-\lambda_a/\lambda_b$ is > 0 of order $1/n$. A sufficient condition for a positive bias is then

$P(\mathbf{Z}_n^T(\mathbf{v}_1^{\otimes 2} - \mathbf{v}_2^{\otimes 2})\mathbf{Z}_n < 0) \geq 1/2$ or

$$P(|\mathbf{Z}_n^T\mathbf{v}_1| < |\mathbf{Z}_n^T\mathbf{v}_2|) \geq 1/2. \tag{5}$$

The projections $\mathbf{Z}_n^T\mathbf{v}_1, \mathbf{Z}_n^T\mathbf{v}_2$ are uncorrelated with mean zero. When Z is gaussian, they are IID conditionally on S, and (5) holds with equality. In general, however, whether (4) holds, and therefore whether the bias is positive or negative, depends on the distributions of Z and of S. Whether the bias is positive or negative, in turn, determines whether the Type II or Type I error rate of Begg's test is inflated. Fig. 2 presents the results of a small simulation where the data exhibit negative bias.

2.4 Bias in the gaussian model

The study effects are often modeled as gaussian by appealing to the CLT, e.g., in the original paper Begg and Mazumdar (1994). In this situation, the bias takes the form given by Theorem 1.

Theorem 1. *Given IID* $(Y_j, S_j), j = 1, \ldots, n,$ *such that*

$$(Y_1, \ldots, Y_n)|(S_1, \ldots, S_n) \sim N(0, \mathrm{diag}(1/S_1^2, \ldots, 1/S_n^2)),$$

$S > 0$ *a.s.,* S *has a continuous lebesgue density,* $\mathrm{E}(S^2) < \infty$, *and* $P(S^2 \leq s)$ *is* $O(s^\epsilon)$ *for some* $\epsilon > 0$, *then* $\mathrm{Var}(\sqrt{n}\tau) \to 4/9 - \frac{(\mathrm{E}\,|S_1 - S_2|)^2}{\pi\,\mathrm{E}(S^2)}$.

For a general distribution for the response Z, in many typical cases, the limiting variance is

$$4/9 - \frac{4(\mathrm{E}\,|S_1 - S_2|)^2}{E(S^2)}\,\mathrm{E}(f_Z(Z))(2\,\mathrm{E}(ZF_Z(Z)) - \mathrm{E}(f_Z(Z))). \tag{6}$$

This result follows by exploiting the theory of U-processes (Nolan and Pollard 1988). The proof given here for the gaussian case in Theorem 1 uses elementary methods. For gaussian

Z, $\mathrm{E}(f_Z(Z)) = \mathrm{E}(ZF_Z(Z)) = 1/(2\sqrt{\pi})$, which leads back to the special case given in the theorem. Expression (6) also gives a criterion for the direction of the bias, i.e., the sign of $2\,\mathrm{E}(ZF_Z(Z)) - \mathrm{E}(f_Z(Z))$.

Theorem 1 states that when the study effects are gaussian the bias depends on the ratio

$$r = \frac{(\mathrm{E}\,|S_1 - S_2|)^2}{\mathrm{E}(S^2)}$$

of the squared mean absolute difference to the second moment of the distribution of study precisions. The quantity is scale free, but depends on the location of S through the denominator. If the distribution of S can be translated toward 0 while maintaining the requirement that $S \geq 0$, doing so leaves the numerator unchanged and can only decrease denominator. The bias is therefore maximized when the precision of a study may be arbitrarily small.

For S uniformly distributed on $[a, b], a \geq 0$, $r = \frac{(b-a)^2}{3(a^2+b^2+ab)}$, which is maximized at 1/3 when $a = 0$. For the general beta distribution with parameters $a > 0, b > 0$, the form of r is given in Table 1. A continued fraction approximation suggests it is maximized at $(a, b) \approx (.15, .39)$ with value $\approx .66$, an asymmetric, bimodal density .

For S exponentially distribution, $r = 1/2$. For a gamma distribution with shape parameter a, $r = \frac{16a}{a+1}(B(1/2, a, a+1) - 1/2)^2$, where $B(1/2, k, k+1) = \int_0^{1/2} x^k(1-x)^{k+1}dx$ is an incomplete beta integral. Numerical evaluation suggests this expression is maximized at $k \approx .54$ with value $\approx .56$.

When S has a pareto distribution with shape parameter a, i.e., $f_S(s) = a\sigma^a/s^{a+1}\{s > \sigma\}$ when the scale parameter is σ,

$$r = \frac{4a(a-2)}{((2a-1)(a-1))^2},$$

defined for $a > 2$ and maximized over a at the unique root of $-2a^3 + 6a^2 - 2a - 1$ on $(2, \infty)$,

$$a = 1 + 2\sqrt{2/3}\cos\left(1/3\arctan\left(\sqrt{101/27}\right)\right) \approx 2.53,$$

where it takes the value $\approx .14$.

In practice, one may attempt to estimate r using the data. This approach may introduce bias of its own, particularly when using common ratio estimators in conjunction with smaller meta-analyses. The performance of an example of this approach is presented in Section 3.

3 Simulations

We consider the Types I and II error rates for the standard and debiased versions of Begg's test. To carry out the debiased version of Begg's test, we used both the true asymptotic bias, which depends on knowing the distribution of S, and an approximation based on the data. For the latter, given a sample of study precisions $s_1, \ldots, s_n$, $r = (\mathrm{E}\,|S_1 - S_2|)^2/E(S^2)$ was estimated by

$$\left(\binom{n}{2}^{-1} \sum_{j<k} |s_j - s_k|\right)^2 \Big/ \overline{s^2}.$$

The scripts used to run the simulations and produce the figures in this section, as well as a supporting `R` package, are available at `https://github.com/haben-michael/begg-public`.

1. FPR control

   To examine the Type 1 error rate, data is generated under the null model (2.3). The distributions of S considered were uniform, beta, exponential, gamma, and pareto, with the parameters given in Section 2.4. Three meta-analysis sample sizes were considered, 25, 75, and 150. The sizes 25 and 75 were chosen to match the sizes used in Begg and Mazumdar (1994). The size 150 was chosen to illustrate the conclusion of Theorem 1 that the bias in Begg's test persists with large sample sizes. The test was conducted at a nominal level of 5%. There were 1000 monte carlo repetitions.

   The results are presented in Table 2. The unadjusted Begg test is conservative and consistent with the discussion in Section 2.2 this bias does not go away with increased sample size. The magnitude of the asymptotic bias follows the order one expects from Table 1: i.e., beta, gamma, uniform, pareto from most to least severe. For the small meta-analyses the corrected test exceeds the nominal level by 1–3%, matching the nominal level for the larger meta-analysis. There does not appear to be much loss in approximating the correct variance using the data rather than using the true variance, even for the smaller meta-analysis.

Next, we examine the Type 1 error rate under the normal random effects model, commonly used to model effect heterogeneity (DerSimonian and Laird 1986). Under this model, the variances of the study effects, denoted σ^2 in (1), are not the variances reported by study authors. Instead, the reported variances represent only the within-study variance, which is inflated by a between-study variance τ^2 common to the studies to obtain the marginal variances of the study effects, i.e., $Y_j \sim \mathcal{N}(\mu, \sigma^2 + \tau^2), j = 1, \ldots, n$. Application of Theorem 1, which uses the marginal variance, therefore requires estimation of the between-study variance, for which we use a standard method of moments estimator (DerSimonian and Laird 1986).

Table 3 presents the observed Type 1 error rates of the unadjusted Begg test using the true between-study variance, the adjusted Begg test also using the true between-study variance, and adjusted Begg test using the estimated between-study variance. The results are similar to those observed in Table 2. Also presented for each within-study precision distribution is the average I^2, a standard measure of effect heterogeneity.

2. Type II error rate

We examine the Type II error rate under an alternative considered by Begg and Mazumdar (1994). Under this alternative, a study with p-value p is selected for publication with probability $\propto \exp(-bp)$, with $b \geq 0$. The parameter b controls the strength of selection, with $b = 0$ corresponding to no selection. The choice of selection function was informed by studies of selection bias contemporaneous with Begg and Mazumdar

(1994). The distributions of S considered for this simulation were uniform, beta, exponential and gamma, and the sizes of the meta-analyses considered were 25, 75, and 150.

Power curves are presented in Fig. 3. For ease of interpretation, the alternatives are parameterized by the proportion of studies selected, rather than b. The improvement in power across distributions and sizes has a median value of 17%. The estimator based on the approximation to the true asymptotic variance performs similarly to the oracle estimator.

4 Data analysis

As an application, we describe three meta-analyses chosen to illustrate different conclusions drawn by the standard and bias-corrected Begg's test. In the typical situation that Begg's test reports a p-value far from the analyst's chosen threshold, the bias-corrected test will agree with the biased test. In the selected examples, the standard Begg's test reports p-values in the range $5-10\%$ and would be insignificant at the 5% level. The three examples were chosen to contrast the stength of evidence for publication bias conveyed by a funnel plot, the conventional informal test for publication bias. In the first, the authors see little evidence for publication bias based on a funnel plot. In the second, the authors are unable to determine the risk of publication bias. In the third, the authors caution that studies have likely been omitted.

In a 2005 analysis, Van de Laar et al. (2005) assess the therapeutic effects of alpha-glucosidase inhibitors in treating type 2 diabetes mellitus. As part of this analysis, they examine the change in body weight under treatment. In a meta-analysis based on 13 randomized trials of at least 12 weeks' duration together involving 864 subjects, the authors find little or no effect of the treatment on weight, contrary to expectations based on earlier work. The authors assess the likelihood of publication bias as low based on a funnel plot (Fig. 4), and Begg's test gives a p-value of 7.4% . The bias-corrected test, however, rejects at the 5% level with a p-value of 2.7%. In this case the possibility of publication bias, suggested by the bias-corrected test, is arguably of less concern as the subsequent meta-analysis can't be a false positive, failing to be significant anyway.

In a 2016 analysis, McNicol et al. (2016) assess the therapeutic effects of intravenously administered paracetamol in treating postoperative pain. As part of their analysis, they examine the reduction in opioids administered under treatment by paracetamol. In a meta-analysis of 13 randomized trials together involving 777 subjects, the authors find a highly significant reduction of 1.92 mg, with a 95% CI $(-2.41, -1.42)$. The authors assess the quality of the data as moderate, noting that the risk of selective reporting is unclear. In this case, a standard Begg's test gives a p-value 5.2%, just insignificant at the 5% level, whereas the bias-corrected test returns a p-value 2.6%, suggesting publication bias is present and casting doubt on the validity of the significant result of the subsequent meta-analysis. A funnel plot (Fig. 4) likewise suggests the presence of publication bias.

In a 2020 analysis, Hooper et al. (2020) assess the effect of low-fat intake on body weight in populations not seeking to lose weight. The authors conduct a meta-analysis on 26 randomized trials of at least six months' duration involving $50,907$ subjects. They find that a low-fat diet is associated with a -1.56 kg difference in weight, with a 95% CI $(-1.88, -1.23)$, and a p-value reported as $< .00001$. Though the trials in this analysis were chosen in part for their low risk of bias, the authors warn of the possibility of selection bias based on a funnel plot (Fig. 4) and other analyses. Begg's test, however, gives a p-value 7.8%, whereas the corrected test gives a p-value 3.3%, consistent with the authors' suspicions.

5 Conclusion

We have examined the causes of a known bias in Begg's test and quantified it in the common model of normally distributed studies. We have also suggested a debiased estimator that matches the nominal variance in the limit as the number of studies grows, unlike the standard estimator, which remains conservative. Although simulations suggest the corrected estimator is somewhat anticonservative on smaller meta-analyses, from the perspective of an analyst concerned primarily about the integrity of the meta-analysis, exceeding the nominal level may be preferable to falling under it, as discussed earlier.

Various avenues of further research suggest themselves. First, the direction of the bias of Begg's test outside of the gaussian model, characterized by (4) for symmetric distributions, has not been fully explored here. It is possible that the direction is the same for certain common large classes, e.g., symmetric unimodal distributions. Second, even in the gaussian model, the magnitude of the bias given by $\frac{(\mathrm{E}\,|S_1-S_2|)^2}{\pi\,\mathrm{E}(S^2)}$ has only here been studied in relation to a few select distributions for the study variance. For example, a distribution maximizing this quantity has not been produced. Finally, a glance at Fig. 3 shows that, even with the boost in power provided by the debiased estimator, the power curves of Begg's test are not very reassuring. There are competitors to Begg's test, such as Egger's test and the more recent test of Lin and Chu (2018). However, quantitative conditions have not been provided to assist an analyst in deciding which test to prefer in given situations.

References

Colin Begg. Publication bias. In Harries Cooper and Larry Hedges, editors, *The Handbook of Research Synthesis*, pages 399–409. Russel Sage Foundation, New York, 1994.

Colin Begg and Madhuchhanda Mazumdar. Operating characteristics of a rank correlation test for publication bias. *Biometrics*, pages 1088–1101, 1994.

Rebecca DerSimonian and Nan Laird. Meta-analysis in clinical trials. *Controlled clinical trials*, 7(3):177–188, 1986.

Miriam Gjerdevik and Ivar Heuch. Improving the error rates of the begg and mazumdar test for publication bias in fixed effects meta-analysis. *BMC Medical Research Methodology*, 14(1):1–16, 2014.

Lee Hooper, Asmaa S Abdelhamid, Oluseyi F Jimoh, Diane Bunn, and C Murray Skeaff. Effects of total fat intake on body fatness in adults. *Cochrane Database of Systematic Reviews*, (6), 2020.

Lifeng Lin and Haitao Chu. Quantifying publication bias in meta-analysis. *Biometrics*, 74 (3):785–794, 2018.

Ewan D McNicol, McKenzie C Ferguson, Simon Haroutounian, Daniel B Carr, and Roman Schumann. Single dose intravenous paracetamol or intravenous propacetamol for postoperative pain. *Cochrane database of systematic reviews*, (5), 2016.

Deborah Nolan and David Pollard. Functional limit theorems for u-processes. *The Annals of Probability*, 16(3):1291–1298, 1988.

Floris A Van de Laar, Peter LBJ Lucassen, Reinier P Akkermans, Eloy H Van de Lisdonk, Guy EHM Rutten, and Chris Van Weel. Alpha-glucosidase inhibitors for type 2 diabetes mellitus. *Cochrane database of systematic reviews*, (2), 2005.

Appendix

Proof of Theorem 1. As discussed above, any bias in the asymptotic variance relative to 4/9 is due to the term

$$P\left(\frac{Z_1 - Z_2}{S_1 - S_2} \wedge \frac{Z_3 - Z_4}{S_3 - S_4} > \hat{\theta}\right) - 1/4. \tag{7}$$

The theorem follows on showing that

$$P\left(\frac{Z_1 - Z_2}{S_1 - S_2} \wedge \frac{Z_3 - Z_4}{S_3 - S_4} > \hat{\theta} \mid S_1, \ldots, S_n\right) - \frac{1}{4} = -\frac{|S_1 - S_2||S_3 - S_4|}{4\pi \sum_{j=5}^n S_j^2} + o(1/\sum_{j=5}^n S_j^2)$$

and that n times the right-hand side is uniformly integrable, since then the LLN gives

$$\begin{aligned}
\mathrm{Var}(\sqrt{n}\tau) - 4/9 &= 4n \cdot (Eq.\ (7)) + o(1) \\
&= 4n\,\mathrm{E}\left(-\frac{|S_1 - S_2||S_3 - S_4|}{4\pi \sum_{j=5}^n S_j^2} + o(1/\sum_{j=1}^n S_j^2)\right) + o(1) \qquad (8) \\
&\to -\,\mathrm{E}\left(\frac{(\mathrm{E}\,|S_1 - S_2|)^2}{\pi\,\mathrm{E}(S^2)}\right).
\end{aligned}$$

Since $Z_1, \ldots Z_n$, are independent of $S_1, \ldots, S_n$, the probability in (7) is

$$\begin{aligned}
&P\left(\frac{\sum_{j=5}^n Z_j S_j}{\sum_j S_j^2} < \left(\frac{Z_1 - Z_2}{S_1 - S_2} - \frac{\sum_{j=1}^4 Z_j S_j}{\sum_j S_j^2}\right) \wedge \left(\frac{Z_3 - Z_4}{S_3 - S_4} - \frac{\sum_{j=1}^4 Z_j S_j}{\sum_j S_j^2}\right) \middle| \mathbf{S}_n\right) \\
&= \mathrm{E}\left(\Phi\left(\frac{\sum_j S_j^2}{\sqrt{\sum_{j=5}^n S_j^2}}\left(\frac{Z_1 - Z_2}{S_1 - S_2} - \frac{\sum_{j=1}^4 Z_j S_j}{\sum_j S_j^2}\right) \wedge \left(\frac{Z_3 - Z_4}{S_3 - S_4} - \frac{\sum_{j=1}^4 Z_j S_j}{\sum_j S_j^2}\right)\right) \middle| \mathbf{S}_n\right) \\
&= \mathrm{E}\left(\Phi(W_0 \wedge W_1) \mid \mathbf{S}_n\right), \qquad (9)
\end{aligned}$$

where W_0, W_1 are jointly normal conditionally on $\mathbf{S}_n = (S_1, \ldots, S_n)$ with mean 0, variances

$$V_0 = \frac{2}{(S_1 - S_2)^2} + \frac{2}{\sum_j S_j^2} + \frac{\sum_1^4 S^2}{(\sum_j S_j^2)^2}, \quad V_1 = \frac{2}{(S_3 - S_4)^2} + \frac{2}{\sum_j S_j^2} + \frac{\sum_1^4 S^2}{(\sum_j S_j^2)^2},$$

and covariance

$$\rho\sqrt{V_0 V_1} = -\frac{2}{\sum_j S_j^2} + \frac{\sum_1^4 S^2}{(\sum_j S_j^2)^2}.$$

The density $f_{W_0 \wedge W_1}$ of the minimum of a bivariate normal pair is readily available and substitution into (9) gives

$$\begin{aligned}\mathrm{E}\left(\Phi(W_0 \wedge W_1) \mid \mathbf{S}_n\right) &= \int_{-\infty}^{\infty} \Phi(u) f_{W_0 \wedge W_1}(u) du \\ &= \sum_{j \in \{0,1\}} \frac{1}{\sqrt{V_j}} \int_{-\infty}^{\infty} \Phi(u) \Phi\left(\frac{u}{\sqrt{1-\rho^2}}\left(\frac{\rho}{\sqrt{V_j}} - \frac{1}{\sqrt{V_{1-j}}}\right)\right) \phi\left(\frac{u}{\sqrt{V_j}}\right) du.\end{aligned}$$

The integral inside the sum has the form

$$\int_{-\infty}^{\infty} \Phi(\alpha_j u) \Phi(\beta_j u) \phi(\gamma_j u) du$$

with $\alpha_j = 1, \beta_j = \frac{1}{\sqrt{1-\rho^2}}\left(\frac{\rho}{\sqrt{V_j}} - \frac{1}{\sqrt{V_{1-j}}}\right)$, and $\gamma_j = \frac{1}{\sqrt{V_j}}$. This integral may be computed by differentiating first with respect to α and then β to put it in terms of elementary functions,

$$\int_{-\infty}^{\infty} u^2 \phi(\alpha_j u) \phi(\beta_j u) \phi(\gamma_j u) du = \frac{1}{2\pi(\alpha_j^2 + \beta_j^2 + \gamma_j^2)^{3/2}},$$

then integrating twice to obtain $\frac{1}{2\pi\gamma_j} \arctan\left(\frac{\alpha_j \beta_j}{\gamma_j \sqrt{\alpha_j^2 + \beta_j^2 + \gamma_j^2}}\right)$ as the required definite integral up to a constant. The constant may be determined by taking limits as $1/(4\gamma_j)$, giving

$$\mathrm{E}\left(\Phi(W_0 \wedge W_1) \mid \mathbf{S}_n\right) = \frac{1}{2} + \frac{1}{2\pi} \sum_j \arctan\left(\frac{\alpha_j \beta_j}{\gamma_j \sqrt{\alpha_j^2 + \beta_j^2 + \gamma_j^2}}\right).$$

Let $\Delta_0 = S_1 - S_2$ and $\Delta_1 = S_3 - S_4$. After simplification, the last expression is

$$\frac{1}{2} + \frac{1}{2\pi} \sum_j \arctan\left(-\frac{1}{\Delta_j^2}\left(\frac{1}{\Delta_0^2 \Delta_1^2} - \left(\frac{1}{\Delta_0^2} + \frac{1}{\Delta_1^2}\right) \frac{1}{2\sum_j S_j^2}\right)^{-1/2}\right).$$

Let $u = 1/\sum_{j=5}^{n} S_j^2$, so by the LLN $u \to 0$ almost surely. When $u = 0$, the last expression is $1/2 + (2\pi)^{-1} \sum_j \arctan\left(-|\frac{\Delta_{1-j}}{\Delta_j}|\right) = 1/4$. Expanding about $u = 0$ and simplifying,

$$\mathrm{E}\left(\Phi(W_0 \wedge W_1) \mid \mathbf{S}_n\right) = \frac{1}{4} + \frac{u}{2\pi} \sum_j \frac{\mathrm{d}}{\mathrm{d}u} \arctan\left(-\frac{1}{\Delta_j^2}\left(\frac{1}{\Delta_0^2\Delta_1^2} - \left(\frac{1}{\Delta_0^2} + \frac{1}{\Delta_1^2}\right)\frac{u}{2}\right)^{-1/2}\right)\Bigg|_{u=0} + o(u)$$

$$= \frac{1}{4} - \frac{u}{4\pi}|\Delta_0\Delta_1| + o(u).$$

Passing the limit into the expectation in 8 follows from the eventual uniform integrability of the sequence $n \cdot u = n / \sum_{j=5}^n S_j^2, n = 1, 2, \ldots$, which in turn follows from Lemma 1.

□

Lemma 1. *If $X_1, X_2, \ldots$ are nonnegative and IID, then the sequence of reciprocals of the sample means $n/(\sum_{j=1}^n X_j), n = n_0, n_0 + 1, \ldots$, is uniformly integrable for some n_0 if and only if the common CDF of the X_j is $O(x^\epsilon)$ for some $\epsilon > 0$.*

Proof. First, $n/(\sum_{j=1}^n X_j)$ has moments > 1, say $1+\epsilon$, for n large enough. As $P(\frac{1}{n}\sum_{j=1}^n X_j < x) \leq P(X_1 < nx)^n$,

$$\begin{aligned}
\mathrm{E}\left(\left(\frac{n}{\sum_{j=1}^n X_j}\right)^{1+\epsilon}\right) &= (1+\epsilon)\int_0^\infty x^\epsilon P\left(\frac{n}{\sum_{j=1}^n X_j} > x\right) dx \\
&\leq (1+\epsilon)\left(1 + \int_1^\infty x^\epsilon P\left(\sum_{j=1}^n X_j < \frac{n}{x}\right) dx\right) \\
&\lesssim (1+\epsilon)\left(1 + \int_1^\infty x^\epsilon \left(\frac{n}{x}\right)^{n\epsilon} dx\right),
\end{aligned}$$

which is finite for $n > 1/\epsilon - 1$. Next, for such n, the sample means $\frac{1}{n}\sum_{j=1}^n X_j, n = 1, 2, \ldots$, are a reverse martingale with respect to $\mathcal{F}_n = \sigma\{\sum_{j=1}^n X_j, \sum_{j=1}^{n+1} X_j, \ldots\}$. The conditional form of Jensen's inequality applied to the convex function $x \mapsto x^{-(1+\epsilon)}$ on $\mathbb{R}^+$ gives, for

$k \in \mathbb{N}$,

$$\mathrm{E}\left(\left(\frac{n+k}{\sum_{j=1}^{n+k} X_j}\right)^{1+\epsilon}\right) = \mathrm{E}\left(\left(\mathrm{E}\left(\frac{1}{n}\sum_{j=1}^{n} X_j \middle| \mathcal{F}_{n+k}\right)\right)^{-(1+\epsilon)}\right)$$
$$\leq \mathrm{E}\left(\left(\frac{1}{n}\sum_{j=1}^{n} X_j\right)^{-(1+\epsilon)}\right) = \mathrm{E}\left(\left(\frac{n}{\sum_{j=1}^{n} X_j}\right)^{1+\epsilon}\right).$$

The reciprocals of the sample means are therefore L^p–bounded with $p = 1+\epsilon$ for all large n, implying that they are uniformly integrable.

Conversely, if $P(X < x)$ isn't $O(x^\epsilon)$ for any ϵ, there are sequences $\epsilon_n \to 0, x_n \to 0, x_n < 1$, such that $P(X < x_n) > x_n^{\epsilon_n}$. Then as $P(\frac{1}{m}\sum_{j=1}^{m} X_j < x_n) > P(X < x_n)^m > x_n^{m\epsilon_n}$,

$$\begin{aligned}
\mathrm{E}\left(\frac{m}{\sum X_j}\right) &= \int_0^\infty P\left(\frac{m}{\sum X_j} > x\right) dx \\
&\geq \int_1^\infty P\left(\frac{1}{m}\sum_j X < 1/x\right) dx \\
&\geq \sum_{j=1}^{\infty} x_j^{m\epsilon_j}(1/x_{j+1} - 1/x_j) \\
&\geq \sum_{j=j_0}^{\infty}(x_j/x_{j+1} - 1),
\end{aligned}$$

with j_0 chosen so that $m\epsilon_j < 1$ when $j \geq j_0$. The condition $x_n \to 0$ then implies

$$\begin{aligned}
\sum_{j=j_0}^{\infty}(x_j/x_{j+1} - 1) &\geq \sum_{j=j_0}^{\infty} \log(x_j/x_{j+1}) \\
&= \log\left(\prod_{j=j_0}^{\infty} x_j/x_{j+1}\right) = \infty,
\end{aligned}$$

so the reciprocals of the sample means aren't integrable. $\square$

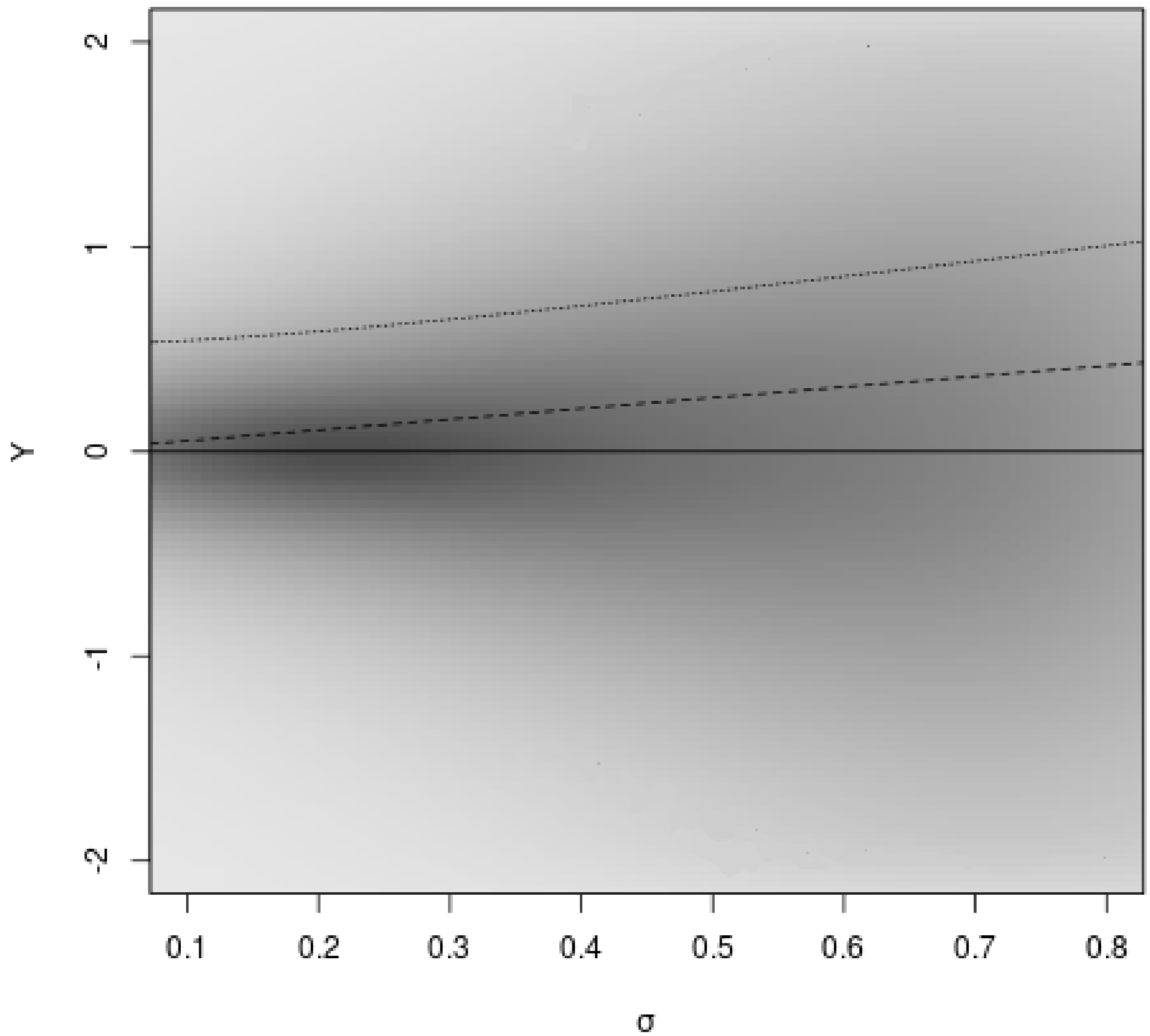

Figure 1: The effect of a simple hard thresholding selection model on the study means. Overlaid on the joint density of Y and σ, indicated by grayscale, is the mean $\theta = 0$ of $Y|\sigma$ before selection and exhibiting no trend (solid line), a threshold line corresponding to rejecting studies with a one-sided p-value $> .3$ (dashed line), and the mean of the studies after selection, exhibiting a trend that Begg's test can pick up (dotted line).

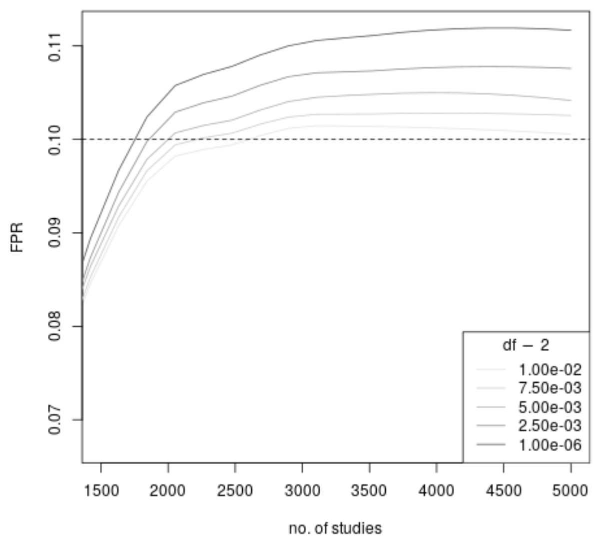

Figure 2: The observed Type I error rate of Begg's test compared to a nominal rate of .1 when the distribution of the response Z is based on a standardized Student's t distribution, i.e., $Z = T/\sqrt{df/(df-2)}$ where T follows a Student's t with df degrees of freedom. The error rate slightly exceeds the nominal rate, particularly as df decreases to the boundary case at 2. In contrast, when Z is gaussian, the error rate falls below the nominal rate.

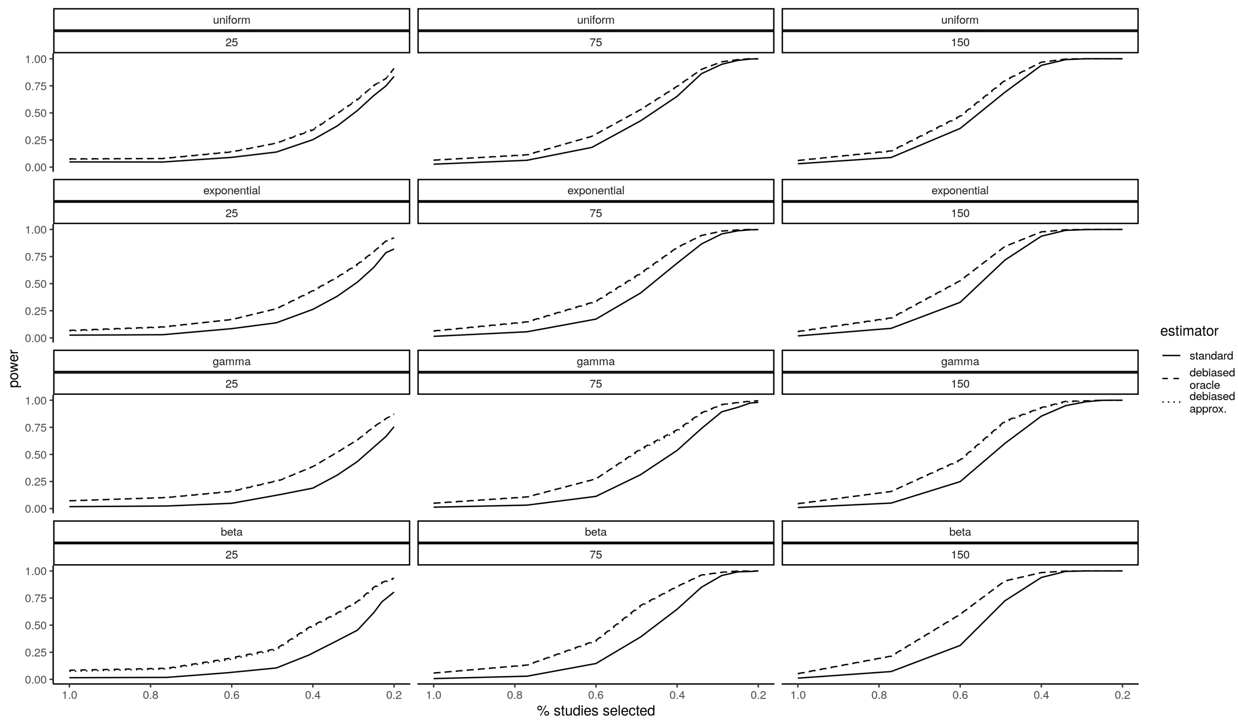

Figure 3: Power curves of standard Begg's test, after debiasing using true asymptotic bias, and after debiasing using estimated asymptotic bias. The alternatives are parameterized by the proportion of studies selected. The estimator debiased using an estimate of the true bias and the true bias itself overlap.

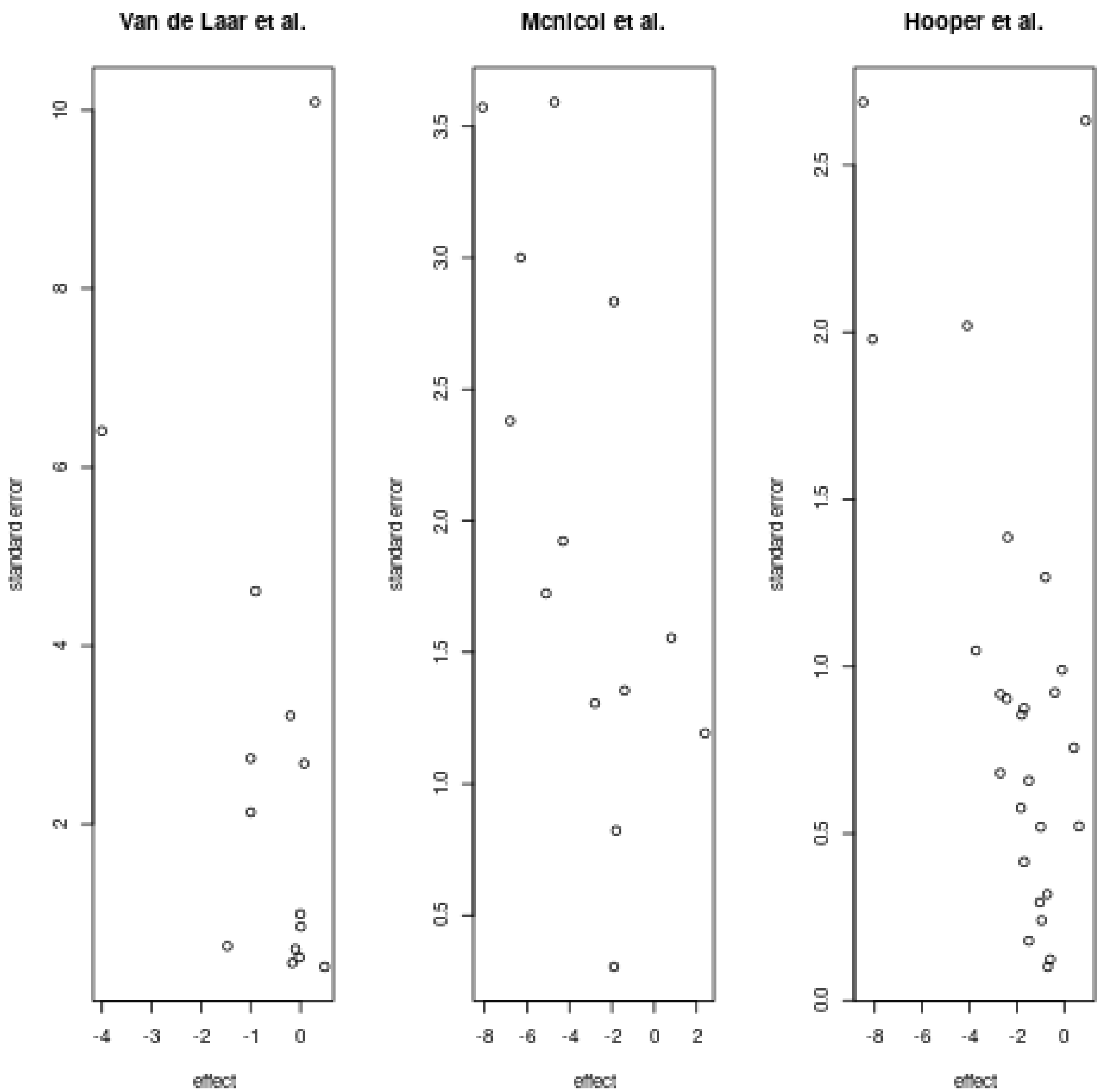

Figure 4: Funnel plots of the three meta-analyses described in Section 4. These were interpreted as suggesting lower, moderate, and higher possibilities of publication bias, going from left to right.

distribution of S	r	max asymptotic bias
uniform	$\frac{(b-a)^2}{3(a^2+b^2+ab)}$	.11
beta	$\left(\frac{4B(a+b,a+b)}{B(a,a)B(b+b)}\right)^2 \frac{a+b+1}{a(a^2+ab+a+b)}$	.21
gamma	$\frac{16a}{a+1}(B(1/2,a,a+1)-1/2)^2$	.18
pareto	$\frac{4a(a-2)}{((2a-1)(a-1))^2}$	.04

Table 1: Shape families of some nonnegative RVs, $r = \text{bias} \times \pi = (\mathrm{E}\,|S_1 - S_2|)^2/E(S^2)$ in terms of the shape parameters, and the maximum bias of the asymptotic variance over the shape family. The maximum variance was obtained numerically for the beta and gamma distributions.

	meta-analysis size		
precision distribution	25	75	150
uniform	0.04, 0.07, 0.07	0.03, 0.05, 0.05	0.03, 0.05, 0.05
exponential	0.03, 0.08, 0.08	0.02, 0.06, 0.06	0.02, 0.05, 0.05
gamma	0.02, 0.07, 0.07	0.02, 0.06, 0.06	0.01, 0.05, 0.05
beta	0.01, 0.08, 0.07	0.01, 0.06, 0.05	0.01, 0.05, 0.05
pareto	0.05, 0.06, 0.06	0.04, 0.06, 0.06	0.04, 0.05, 0.05

Table 2: False positive rates for the standard Begg's test, after debiasing using the true asymptotic bias, and after debiasing using an estimate of the asymptotic bias.

	meta-analysis size		
precision distribution	25	75	150
uniform ($I^2 = 0.54$)	0.05, 0.07, 0.07	0.03, 0.05, 0.05	0.03, 0.05, 0.05
exponential ($I^2 = 0.85$)	0.05, 0.07, 0.07	0.04, 0.06, 0.06	0.04, 0.05, 0.05
gamma ($I^2 = 0.67$)	0.03, 0.06, 0.06	0.02, 0.06, 0.06	0.02, 0.05, 0.05
beta ($I^2 = 0.39$)	0.01, 0.07, 0.07	0.01, 0.05, 0.05	0.01, 0.06, 0.06
pareto ($I^2 = 0.92$)	0.06, 0.06, 0.06	0.06, 0.06, 0.06	0.06, 0.06, 0.06

Table 3: False positive rates under effect heterogeneity for the standard Begg's test using the true (unobserved) between-study variance, after debiasing using an estimate of the asymptotic bias while still using the true (unobserved) between-study variance, and after debiasing using an estimate of the asymptotic bias and an estimate of the between-study variance.