FPR of the Egger and Skew Tests for Publication Bias

Haben Michael[1]

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

SUMMARY: Publication bias undermines the results of meta-analyses and other systematic reviews. Applying a formal test for publication bias is therefore part of many protocols for conducting a meta-analysis. Concerns have been raised over the error rates of these tests, particularly since the small samples often encountered in meta-analyses may not justify the asymptotic assumptions underlying these tests. this makes it sound like we'll be doing a finite sample analysisWe examine the FPR of Egger's test and the residual skew test of Lin and Chu (2018) in three models of increasing generality. We find that while Egger's test may outperform the skew test in the more restrictive models, it has significant drawbacks in the more general and realistic model.
KEYWORDS: EDGEWORTH EXPANSIONS, META-ANALYSIS, PUBLICATION BIAS.

1 Introduction

When an area of scientific inquiry has produced a body of results on a similar topic, meta-analysis is a technique for combining them to produce a single, hopefully more certain conclusion. A threshold assumption of the technique is that the included studies represent all the studies conducted on the inquiry, rather than, say, just those studies that produced a statistically significant outcome. Several methods, formal and informal, have been proposed to test the assumption that the studies are representative.

Egger's test (Egger et al., 1997) is one of the earliest and perhaps most popular of the formal tests. Let θ denote the true value of a quantity under study, and suppose each of n studies produces an estimate of the quantity and an estimate of its standard error,

$$(Y_1, \hat{\sigma}_1), \dots, (Y_n, \hat{\sigma}_n), \qquad \mathrm{E}(Y_i) \approx \theta. \tag{1}$$

Egger's test statistic is the t-statistic for the intercept

$$T_n^{(E)} = \hat{\beta}_0 / \text{s.e.}(\hat{\beta}_0) \tag{2}$$

in the simple linear regression

$$Y_i/\hat{\sigma}_i \text{ vs. intercept and } 1/\hat{\sigma}_i,\ i = 1, \dots, n. \tag{3}$$

The premise is that the regressors $1/\hat{\sigma}_i$ are proxies for study quality and other factors that may improperly influence a decision to publish, and that in the absence of publication bias the regression line should nearly run through the origin with a slope near the effect θ; see Section 3 for further intuition.

To boost the power of a test, it is natural to look for another, independent stream of information that bears on the distinction between the null and the alternative hypotheses. Since the test statistic (2) is the fitted coefficient in a linear regression, perhaps the most immediate source of independent information is the regression residuals, and it appears the residuals can effectively distinguish the null from the alternative in common models of publication bias (Lin and Chu, 2018); see also Fig. 1. Lin and Chu (2018) accordingly proposes to test for publication bias using an estimate of the standardized skew of the observed regression residuals $\hat{\epsilon}_1, \ldots, \hat{\epsilon}_n$,

$$\begin{aligned} T_n^{(s)} &= \sqrt{\frac{n}{6}} \frac{\frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i^3}{(\frac{1}{n-1}\sum_{i=1}^n \hat{\epsilon}_i^2)^{3/2}} \\ &= \frac{(n-1)^{3/2}}{n\sqrt{6}} \frac{\hat{\theta}_3}{\hat{\theta}_2^{3/2}}, \qquad \hat{\theta}_p = \frac{1}{n}\sum_{i=1}^n \hat{\epsilon}^p. \end{aligned} \tag{4}$$

We refer here to this test as the "skew test." The authors show by extensive simulation that this test has better power than competing tests, including Egger's. Power can be further boosted by combining the skew test with the Egger's test, which is stochastically independent under the Gaussian model discussed in Section 2.

The primary purpose of this paper is to investigate whether this power boost comes at a cost to the level of the test. To do so, we investigate the false positive rates of both Egger's test and the skew test. A frequent criticism of formal tests for publication bias is that their assumptions are unlikely to hold, even approximately, at the sample sizes encountered in meta-analyses. We therefore consider three models for meta-analyses of increasing generality. We find that Egger's test may outperform the skew test in the more restrictive models, while the skew test has better robustness properties that recommend its use in the more general, arguably more realistic model. A theme will be the interplay between the number of studies included in the meta-analysis relative to the sizes of the individual studies.

remark: limits of testing skew. result for distributions with exponential shapes.

2 Gaussian and other symmetric models

The most restrictive model we consider is the Gaussian. It is also the only one we consider in which both the Egger and skew tests attain the nominal level as the number of studies grows while the sizes of the studies themselves are of constant order. In fact, both tests approach the nominal level faster than the parametric rate, since the Gaussian distribution's skew-free property carries over to the test statistics, as discussed further below. We present second-order expansions to quantify and compare the convergence.

The Gaussian model of meta-analysis is that, in the absence of publication bias, the study means are conditionally Gaussian with a common mean and variances exactly as reported,

$$(Y_1, \ldots, Y_n) | (\hat{\sigma}_1, \ldots, \hat{\sigma}_n) \sim N(\theta, \mathrm{diag}(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_n^2)). \tag{5}$$

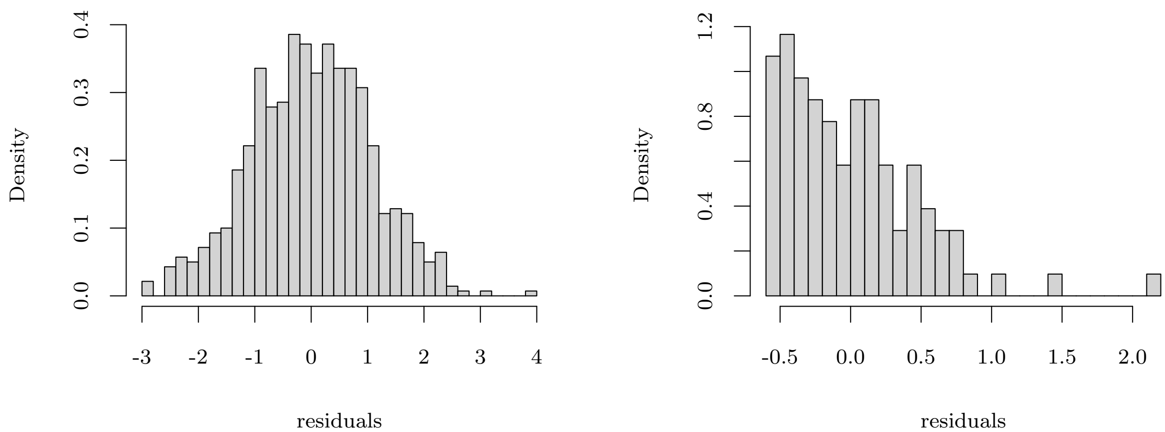


Figure 1: Section 1 (Introduction): The impact of selection on the residuals of an Egger regression under a simple p-value hard thresholding model. On the left are the residuals with no selection, and on the right, the residuals after omitting all studies with p-values exceeding 30%. Following Lin and Chu (2018), the studies are modeled as normal with SD σ, with σ modeled as uniform on the interval $(1, 4)$. The pre-selection population depicted on the left consists of about 700 studies, and the post-selection population consists of about 100 studies.

Many of the common tests for publication bias (Begg and Mazumdar, 1994; Egger et al., 1997; Lin and Chu, 2018) were introduced in the context of this model, sometimes as an explicit assumption and other times as motivation. The Gaussian model also underlies many techniques used in conjunction with meta-analysis, such as Cochran's Q test for the presence of inter-study heterogeneity and the DerSimonian-Laird estimator of inter-study heterogeneity (Cooper et al., 2019). The justification is that the data, being the outcomes of studies, are often averages subject to the CLT. The Gaussian assumption may therefore be regarded as an implicit asymptotic assumption, i.e., that the sizes of the primary studies are large, in addition to the asymptotic regime under which Egger's and the skew test were introduced, i.e., that the number of studies n is large. The double asymptotics are considered in Section 4.

Under the Gaussian model (5), both tests attain the nominal level as the number of primary studies approaches infinity. Moreover, this convergence occurs at an $O(1/n)$ rate in the number of primary studies n. Proposition 1 quantifies the tests' convergence to the Gaussian CDF under the Gaussian model. Let $T_n^{(E)}$ and $T_n^{(s)}$ denote the Egger and skew test statistics based on a sample of n primary studies, as in (1). Let R denote a random variable distributed as the reciprocal standard errors, $R_i \sim 1/\hat{\sigma}_i, i = 1, \ldots, n$.

Proposition 1. *Assuming model* (5) *holds and that* $\mathrm{P}(R^4) < \infty$*,*

$$\mathrm{P}(|T_n^{(E)}| > z_{1-\alpha/2}) = \alpha + \frac{\phi(z_{1-\alpha/2})}{2n}(z_{1-\alpha/2}^3 - z_{1-\alpha/2}) + o(1/n)$$
$$\mathrm{P}(|T_n^{(s)}| > z_{1-\alpha/2}) = \alpha + \frac{\phi(z_{1-\alpha/2})}{n}(\frac{5}{3}z_{1-\alpha/2}^3 - 14z_{1-\alpha/2}) + o(1/n)$$

Remarks on Proposition 1:

1. Since the CDFs converge at an $O(1/n)$ rate, so also do p-values and CIs. This fast convergence mitigates the small-sample problems often encountered in meta-analyses, provided the Gaussian model holds.

2. The CDFs, up to second order, do not depend on the distribution of the reported standard errors $1/\hat{\sigma}_i, 1 \leq i \leq n$. The Egger statistic under the Gaussian model follows a Student's t distribution so it is in fact ancillary with respect to the design matrix, which is where the variances enter into the regression (3).

3. The $O(1/n)$ rate given in Proposition 1 holds for any symmetric distribution, i.e., when $Y_i|\hat{\sigma}_i, i = 1, \ldots, n$, are independent and each follows a symmetric distribution with mean θ and variance $\hat{\sigma}_i^2$. The test statistics are then symmetric, being odd functions of the vector of effects $(Y_1, \ldots, Y_n)$, so the $O(1/\sqrt{n})$ terms in the expansions, corresponding to the skew of the test statistics, vanish.

These points are in part offered to show the unrealistic strength of the Gaussian assumption. For example, it is natural and common to pay attention to the distribution of variances when examining a funnel plot (Rothstein et al., 2005), which should not matter under the Gaussian model by the remarks above.

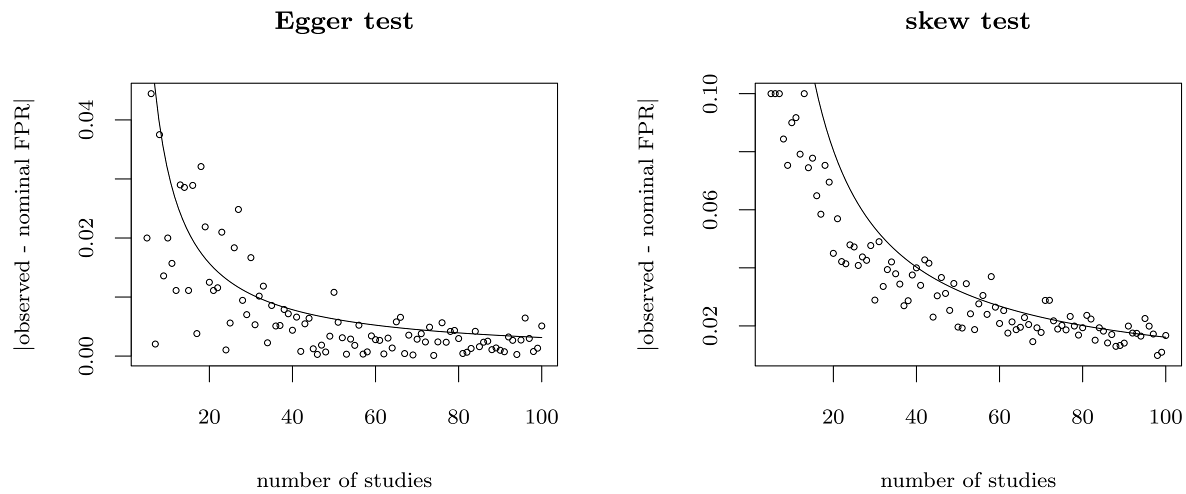


Figure 2: Section 2 (Gaussian model): The false positive rates of both Egger's and the skew test converge to the nominal rate ($\alpha = .1$) at a $1/n$ rate. The circles are differences between the observed and nominal FPR and the solid lines are $O(1/n)$ the differences implied by the second-order Edgeworth expansions given in Proposition 1.maybe put this on a log=log scale))

We illustrate the proposition in a small simulation approximating the difference of the observed and nominal false positive rates of the tests as the number of primary studies n included in the meta-analysis increases. Since the primary studies are Gaussian the only adjustable statistical parameter is the distribution for the study variances. We take them to be uniform on $(1, 4)$ following Lin and Chu (2018), though by the foregoing remarks we don't expect the choice to have an impact in the Gaussian model. Figure 2 plots the difference between the observed and nominal FPRs for a range of sample sizes. Also plotted is the theoretical difference between the observed and nominal FPRs implied by Proposition 1. This theoretical difference is $O(1/n)$, and matches well the observed difference.

REMARK: maybe add remark on pitman efficiency, infinite for skew test.

3 Moment model

In this section we relax the Gaussian assumption and the assumption that the primary studies be identically distributed. While the Egger test remains valid in this setting, the skew test does not, the latter tending to reject whenever the average skew of the standardized study means is non-zero. In the case of the skew test, we will find that the two asymptotic regimes, the number of primary studies and the sample sizes of the primary studies, lie in a certain tension.

The moment model captures the basic principles of meta-analysis. The data consist of primary study statistics and their variances. The primary studies are all taken to model the same underlying quantity θ, and the reported variances are taken to be perfectly accurate, conditionally:

$$\begin{aligned}(Y_1, \hat{\sigma}_1^2), \ldots, (Y_n, \hat{\sigma}_n^2) &\text{ independent}\\ \mathrm{E}(Y_1, \ldots, Y_n \mid \hat{\sigma}_1^2, \ldots, \hat{\sigma}_n^2) &= \theta \mathbb{1}_n \\ \mathrm{Var}(Y_1, \ldots, Y_n \mid \hat{\sigma}_1^2, \ldots, \hat{\sigma}_n^2) &= \mathrm{diag}(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_n^2).\end{aligned} \tag{6}$$

An example of its application is in Section 4 of Lin and Chu (2018).

The assumption that the studies have a common mean can be relaxed somewhat by working in the "random effects model." In this model the means $\theta_i, i = 1, \ldots, n$, of the primary studies are modeled as random with common mean θ and variance τ^2. Marginally, the data resemble the fixed-effects model (6), with $\hat{\sigma}^2 + \tau^2$ in place of σ^2. The difference in practice is that the variance component τ^2 must be estimated at the meta-analysis stage. Since this is a type of estimation error in the variance and we consider more generally the effect of the error in estimating the variances in Section 4, we simply focus on the fixed effects model here.

Under model (6), $E(Y_j/\hat{\sigma}_j \mid \hat{\sigma}_j) = \theta/\hat{\sigma}_j$ and $\mathrm{Var}(Y_j/\hat{\sigma}_j) = 1$. Therefore,

$$y_j/\hat{\sigma}_j = \beta_0 + \beta_1/\hat{\sigma}_j + \epsilon \tag{7}$$

is a correctly specified linear model with independent homoskedastic errors ϵ and $\beta_0 = 0$. The Egger test statistic, i.e. the t-statistic for β_0, is therefore a valid basis for testing the null model (6).

By contrast, the skew test is not level α, even asymptotically, in model (6). As Lin and Chu (2018) point out, the shape of the fitted residuals usually approximates the shape of the centered and standardized primary study means. Proposition 2 shows that the skew test is, up to first order, a test of skew on the standardized effect sizes $Z = (Y - \theta)/\hat{\sigma}$.

Proposition 2. *In addition to the assumptions* (6)*, suppose that the averages of the reported reciprocal standard errors and variances estimates* $\frac{1}{n}\sum_{i=1}^n 1/\hat{\sigma}_i$ *and* $\frac{1}{n}\sum_{i=1}^n 1/\hat{\sigma}_i^2$ *have nonzero probability limits. Suppose also that the* p^{th} *sample moment of the residuals* $\frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i^p$ *in the regression* (7) *has a nonzero probability limit, denoted* θ_p*, for* $p = 2$ *and* 3*. Then*

$$\sqrt{\frac{6}{n}} T_n^{(s)} = \theta_2^{-3/2} \left(\frac{1}{n} \sum_{i=1}^n (\frac{Y_i - \theta}{\hat{\sigma}_i})^3 + 3(\theta \overline{1/\hat{\sigma}} - \overline{Y/\hat{\sigma}}) \right) - 3/2\theta_2^{-5/2}\theta_3 \left(\frac{1}{n} \sum_{i=1}^n (\frac{Y_i - \theta}{\hat{\sigma}_i})^2 - \theta_2 \right) + o_P(1/\sqrt{n}). \tag{8}$$

As $n \to \infty$ all terms in (8) tend to 0 under model (6) other than

$$\frac{1}{n\theta_2^{3/2}} \sum_{i=1}^n (\frac{Y_i - \theta}{\hat{\sigma}_i})^3,$$

proportional to the sample skew of the primary study means. The population skew to which this average tends is zero for the Gaussian or other symmetric distributions. While it may be plausible that the primary study means are each asymptotically Gaussian, there is perhaps little reason to expect them to be Gaussian or even skew-free in the finite samples actually

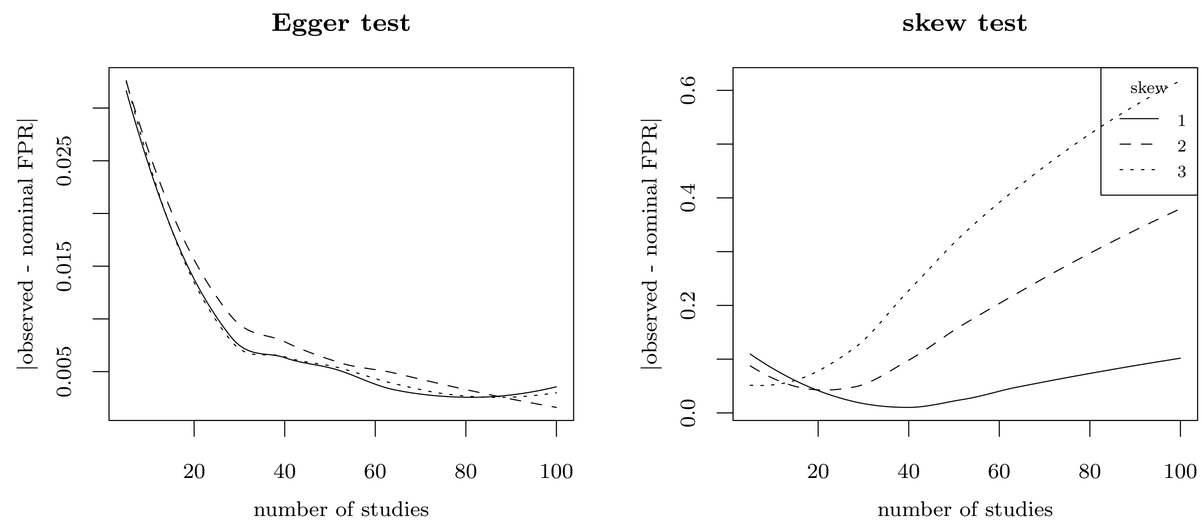


Figure 3: Section 3 (Moment model with log-normal data): As the number of studies grows, while the sizes of the studies remain at the same fixed magnitude, Egger's test is valid even when the data is skewed, but the skew test is not.

used, however large. This non-zero skew can be picked up by the skew test, which will then flag publication bias even in its absence. A larger number of primary studies, without a commensurate increase in the sample sizes of the primary studies, will only exacerbate the problem as the test gains power to detect any non-zero skew, however small.

Meta-analyses with systematically skewed, continuous data may be found in subject areas measuring, e.g., chemical concentrations or ratios; see Higgins et al. (2008) for further examples. For count data, see, e.g., Macaskill et al. (2001). For example, the data may be log-normally distributed (Higgins et al., 2008). We again use a small simulation to demonstrate our theoretical arguments regarding the FPRs of the two tests. See Fig. 3. As anticipated, the presence of skew breaks the skew test's FPR control, while Egger's test appears valid. Also as expected, for smaller skew, the skew test may initially miss the skew but eventually picks it up for larger meta-analyses, resulting in a "U" shape for the error curve.

The proposition and simulation suggest that the poor performance of the skew test relative to Egger's test is tied directly to the asymptotic regime under which meta-analysis is typically analyzed. In this regime, the study sizes remain of the same order while the number of studies grows. In the next section we consider a more general asymptotic regime.

4 Misspecified model

The first of two goals of this section is to relax some of the stronger assumptions in the previous two models in order to analyze the behavior of Egger's test and the skew test in more realistic settings. Mainly, we acknowledge that the regressors $1/\hat{\sigma}_i, i = 1, \ldots, n$, in (3) are in practice estimated from the data. When this measurement error is present the conditional moment assumptions (6) need not hold. From the standpoint of the linear regression (3) underlying Egger's test and the skew test, the result is that heteroscedasticity and, more importantly, endogeneity, may invalidate the test statistics.

Since the study variances are estimated by the primary study authors, allowing for estimation error invites us to model the reported primary study statistics $(Y_i, \hat{\sigma}_i^2)$ in (1). Accordingly, a second goal of this section is to consider asymptotics not only in the number of primary studies but also the sizes of the samples on which the primary studies base the reported effect estimators and variances. In so doing we make explicit the assumption mentioned in Section 2 and lurking in the background of many of the foundational publication bias papers. Under this more general asymptotic regime, we will see that, contrary to the conclusions of Section 3, the skew test may be regarded as asymptotically valid, when Egger's test remains inconsistent.

As above, let $(Y_1, \hat{\sigma}_1), \ldots, (Y_n, \hat{\sigma}_n)$, denote n independent random pairs representing the effects and their standard errors reported by n primary studies. However, we no longer assume $\hat{\sigma}_i$ perfectly represents the SD of Y_i, that the Y_i have a common mean, or that these pairs are identically distributed. At the same time, we assume the data describe a "jagged array" of independent observations, representing the observations on which $(Y_i, \hat{\sigma}_i), i = 1, \ldots, n$, are based. For example, in a meta-analysis based on matched trials the data underlying the primary studies are the differences for each unit, say,

$$\begin{array}{rlllll}
\text{study 1 data:} & W_{11} & W_{12} & \ldots & W_{1,m_1} & \\
\text{study 2 data:} & W_{21} & \ldots & W_{2,m_2} & & \\
\vdots & \vdots & & & & \\
\text{study } n \text{ data:} & W_{n1} & W_{n2} & \ldots & \ldots & W_{n,m_n}
\end{array}$$

whereas in a binary outcome trial the underlying data are event and treatment indicators, say

$$\begin{array}{rlllll}
\text{study 1 data:} & (T_{11}, E_{11}) & (T_{12}, E_{12}) & \ldots & (T_{1,m_1}, E_{1,m_1}) & \\
\text{study 2 data:} & (T_{21}, E_{21}) & \ldots & (T_{2,m_2}, E_{2,m_2}) & & \\
\vdots & \vdots & & & & \\
\text{study } n \text{ data:} & (T_{n1}, E_{n1}) & (T_{n2}, E_{n2}) & \ldots & \ldots & (T_{n,m_n}, E_{n,m_n}).
\end{array} \tag{9}$$

The study effects Y_i and SEs $\hat{\sigma}_i$, $i = 1, \ldots, n$, input into the meta-analsyis are assumed to be smooth functions of averages of these observations. For example, the odds ratio and its standard error are given as such by (14).

Proposition 3 gives the first-order asymptotic behavior of the Egger and skew test statistics under this asymptotic regime, with an emphasis on their biases under an IID design. Let $R_i = 1/\hat{\sigma}_i$ denote the regressors in (3). Let $Z_i = Y_i R_i = Y_i/\hat{\sigma}_i$, again without assuming $\hat{\sigma}^2 = \mathrm{Var}(Y_i)$. Let $\mathrm{P}_n, \mathrm{E}_n$, etc. denote probability, expectation, etc. with respect to the empirical measure on the pairs $(R_i, Y_i), i = 1, \ldots, n,$, e.g., $\mathrm{P}_n\, R^2 = \frac{1}{n}\sum_{i=1}^n R_i^2$.

Proposition 3. *Assuming:*

1. $(R_i, Y_i), i = 1, \ldots, n$, *are smooth functions of averages of* m_i *independent observations*
2. $\sup_i E|R_i^u Y_i^v|^{2+\delta} < \infty, 0 \leq v \leq u, 0 \leq u \leq 3$

3. $\sup_n \mathrm{P}_n R^2 < \infty$

4. $\sup_n \mathrm{Corr}_n(Z, R) < 1$

Then Egger's test statistic may be written

$$T^{(E)}_{n,\{m_i\}}/\sqrt{n} = \mathrm{Var}(T^{(E)}_{n,\{m_i\}}/\sqrt{n})U_1 - \mathrm{P}_n \frac{\mathrm{Corr}(R, Y)}{\sqrt{1 - \mathrm{Corr}(R_i, Y_i)^2}} + O(\mathrm{E}_n(\mathrm{E}\, R^4 + \mathrm{E}\, Y^4 + \mathrm{E}\, Z^4)) + O_P(1/n),$$

where U_1 is standard normal.

Let

$$\epsilon_i = (Y_i - \mathrm{E}\, Y_i - \beta_1(R_i, Y_i)(R_i - \mathrm{E}\, R_i))/\sqrt{\mathrm{Var}\, R_i(1 - \mathrm{Corr}(R_i, Y_i)^2)}$$

denote the population standardized residual from regressing Y_i onto R_i. Then the skew test statistic may be written

$$T^{(s)}_{n,\{m_i\}}/\sqrt{n} = \mathrm{Var}(T^{(s)}_{n,\{m_i\}}/\sqrt{n})U_2 - \mathrm{P}_n\, \mathrm{E}\, \epsilon^3/\sqrt{6} + O(\mathrm{E}_n(\mathrm{E}\, R^6 + \mathrm{E}\, Y^6 + \mathrm{E}\, Z^6)) + O_P(1/n), \tag{10}$$

where U_2 is standard normal, and $\mathrm{E}\, \epsilon_i^3$ is $O(1/\sqrt{m_i})$.

Remarks on Proposition 3:

1. In both cases the expansion consists of a Gaussian term and a non-random bias. We focus on the latter in what follows, as an order of magnitude difference can separate the bias of Egger's test and the skew test.

   For Egger's test the bias is the average of

   $$\mathrm{Corr}(R_i, Y_i)/\sqrt{1 - \mathrm{Corr}(R_i, Y_i)^2} \tag{11}$$

   measuring the strength of the linear association between Y_i and R_i. This quantity vanishes under the commonly encountered meta-analysis assumption that Y_i is mean independent of $\hat{\sigma}_i$ (see Section 3). If however this quantity doesn't vanish entirely, the bias of Egger's test statistic relative to a standard normal is $O(\sqrt{n})$.

   For the skew test the bias is the average skew of the standardized residual from regressing Y_i on R_i. Under the Proposition's assumption that Y_i and R_i are based on IID means, the same holds of the i^{th} residual ϵ_i, which then has an $O(1/\sqrt{m_i})$ skew. The bias of the skew test statistic is then $\sqrt{n} \cdot \frac{1}{n}\sum_{i=1}^n O(1/\sqrt{m_i})$. As a result, the false positive rate of Egger's test may quickly fall to 0 or jump to 1, whereas false positive rate of the skew test may converge to the nominal rate as long as the sizes of the primary studies are large relative to the number of studies.

2. When the underlying study data is exchangeable across arms (two-arm study) or symmetric (in a matched study) the study's contribution to the Egger or skew test bias usually vanishes. Specifically, for most choices of meta-analysis effect size, and all those considered here, the effect Y is an odd function and the reciprocal SE R an even

function of the same underlying data, implying they are uncorrelated when the underlying data is exchangeable or symmetric. The bias of Egger's test (11) is a measure of correlation, so it then vanishes. For the skew statistic, uncorrelatedness implies the residual ϵ_i defined in the statement of Proposition 3 reduces to $Y_i - \mathrm{E}\, Y_i$, a centered odd function which is therefore skew-free. However, the points where the bias vanishes can be isolated or unstable; see Section 4.2.2.

3. As a caveat, we add that, though the proposition doesn't explicitly assume the data is homogenous, its conclusion is framed to emphasize the asymptotic behavior under homogeneity. The effect of variation between the moments of the (Y_i, R_i) for different i, and the effect of variation in the primary study sample sizes m_i, are relegated to error terms. To include the effect of this variation in the bias substantially complicates the expressions for the bias. Suppose, for example, that the mixed moments of (R_i, Y_i) do not vary with i, but the sample sizes m_i may do so. Then, taking into account the first-order effects of m_i, the bias of Egger's test statistic becomes

$$-(\mathrm{E}_n \sqrt{m} + \mathrm{Cov}_n(m, 1/\sqrt{m}))/(\mathrm{E}_n(m)(1/\mathrm{Corr}_n(R,Y)^2 - 1 - \mathrm{Var}_n(\sqrt{m})(\mathrm{E}_n R)^2 \mathrm{Var}_n Y$$

$$+ c_1 \mathrm{Cov}_n(m, 1/\sqrt{m}) + c_2(\mathrm{E}_n m(\mathrm{E}_n(1/\sqrt{m}))^2 - 1) + c_3 \mathrm{Var}_n(\sqrt{m}) \mathrm{E}_n(1/m)))^{1/2}$$

where c_1, c_2, and c_3, are constants depending on the common mixed moments of (R_i, Y_i) but not m_i. This more complicated expression reveals that, if it happens that the empirical variance of the root of the primary study sizes, $\mathrm{Var}_n(\sqrt{m})$, is of order $\sqrt{m}$, then the bias of Egger's statistic can be neutralized by primary study samples large relative to the number of studies, just as with the skew statistic.maybe test with a sim, put in appendix

4.1 Studies based on continuous data

Primary studies based on continuous data commonly report means. In this section we make the simplifying assumption that the data are drawn from a matched design (Cooper et al., 2019). Suppose primary study $i, 1 \leq i \leq n$, collects m_i IID continuous observations $W_{i1}, \ldots, W_{im_i}$, such as by taking the differences in measurements before and after an intervention. Study i then computes the statistics

$$\begin{aligned} \overline{W_{i\cdot}} &= m_i^{-1} \sum_{j=1}^{m_i} W_{ij} \\ \widehat{\mathrm{Var}}(\overline{W_{i\cdot}}) &= m_i^{-1} \sum_{j=1}^{m_i} (W_{ij} - \overline{W_{i\cdot}})^2 \end{aligned} \tag{12}$$

which are included as $(Y_i, \hat{\sigma}_i), i = 1, \ldots, n$, in the Egger regression (3). Egger's statistic or the skew statistic is then formed as before. Corollary 4 evaluates the formulas of Proposition 3 to describes the asymptotic behavior of the resulting test statistics.

Corollary 4. *Assumptions:*

1. *For each i, $W_{ij}, 1 \leq j \leq m_i$ are independent and identically distributed*

2. $\sup_{i,j} \mathrm{E}\, W_{ij}^4 < \infty$

3. $\sup_n \mathrm{Corr}_n(Z, R) < 1$ *see if this can be replaced by others...maybe bc data is continuous. or at least reexpress in terms of w.*

4. *For each i, the common distribution of the $W_{ij}, 1 \leq j \leq m_i$, has a PDF f_i that is continuous on its support and vanishes at infinity, and these PDFs are uniformly bounded,* $\sup_i |f_i|_\infty < \infty$.

5. $m_i \geq 5$, $i = 1, \ldots, n$

Let $\mu_{k,i} = E(W_i^k)/E(W_i^2)^{k/2}$ denote the k^{th} standardized moments of W_{ij}. Based on data $(Y_i, \hat{\sigma}_i) = (\overline{W}_{i\cdot}, 1/R_i)$ given as in (12), there are $U_1, U_2 \sim N(0,1)$ such that also then need to check formulas for variance. the inid case was only proven for the biases.

$$
\begin{aligned}
T^{(E)}_{n,\{m_i\}}/\sqrt{n} = &\sqrt{\mathrm{P}_n \frac{\mu_4 - 1}{\mu_4 - \mu_3^2 - 1}} \frac{U_1}{\sqrt{n}} + \mathrm{P}_n \frac{\mu_3}{\sqrt{\mu_4 - \mu_3^2 - 1}} \\
&+ O(\mathrm{E}_n(\mathrm{E}\, R^4 + \mathrm{E}\, Y^4 + \mathrm{E}\, Z^4)) + \mathrm{P}_n(1/\sqrt{m}) + O_P(1/n)
\end{aligned}
$$

and

$$
\begin{aligned}
T^{(s)}_{n,\{m_i\}}/\sqrt{n} = &\sqrt{1 + \mathrm{P}_n\, O(1/\sqrt{m})} \frac{U_2}{\sqrt{n}} \\
&+ \mathrm{P}_n \frac{1}{\sqrt{m}} \frac{\mu_3 \left(6\mu_3^4 - 2\left(\mu_4 - 4\right)\left(\mu_4 - 1\right)^2 + 3\mu_3\left(\mu_4 - 1\right)\mu_5 - \mu_3^2\left(\mu_6 + 15\mu_4 - 16\right)\right)}{\sqrt{6}\left((\mu_4 - 1)(\mu_4 - \mu_3^2 - 1)\right)^{3/2}} \\
&+ O(\mathrm{E}_n(\mathrm{E}\, R^6 + \mathrm{E}\, Y^6 + \mathrm{E}\, Z^6)) + \mathrm{P}_n(1/m) + O_P(1/n).
\end{aligned}
$$

Remarks on Corollary 4:

1. The behavior of the bias is as anticipated in the remarks following Proposition 3. Except in the case of skew-free data, the bias terms are non-zero, but the bias of Egger's test is an order of magnitude greater than the skew test. Specifically, when $\mu_3 \neq 0$ in 4, the bias of Egger's test statistic relative to a standard normal is $O(\sqrt{n})$ whereas the bias of the skew statistic is $\sqrt{n} \cdot \frac{1}{n}\sum_{i=1}^n O(1/\sqrt{m_i})$. As a result, the false positive rate of Egger's test will converge at an exponential rate to 0 or 1 according as μ_3 is negative or positive. The false positive rate of the skew test, however, will converge to the nominal rate, even in the presence of skew, as long as n is on average $o(m_i)$.

2. Besides the overall order of the bias, Corollary 4 also relates the magnitude of the bias to the moments of the study observations. As noted, the bias disappears for skew-free data. On the other hand, the magnitude of the denominator is governed by Pearson's inequality for random variables with four moments (Wilkins, 1944), asserting that the kurtosis is bounded below by the squared skewness plus one. The equality

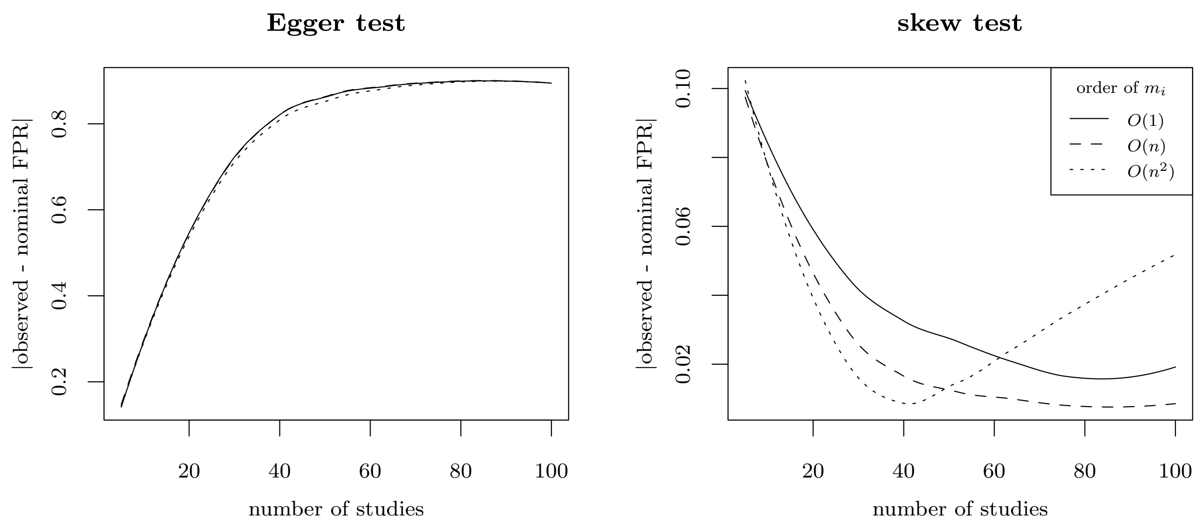


Figure 4: Section 4.1 (Misspecified model with log-normal data): When the study observations have positive skew, as here, the false positive rate of Egger's test rapidly grows to 1. By contrast, the FPR of the skew test vanishes as long as the average study size grows at a faster rate than the number of studies.

case, when the asymptotic biases given in Corollary 4 blows up, occurs for a skewed two-point distribution, which is at the boundary of our continuous model (Corollary 4, Assumption 5). Consistent with this behavior, Michael (2025, Section 3.2) shows that the power of Egger's test also suffers when the study means Y follow skew beta distributions. The asymptotic bias is small when the kurtosis of the primary study observations is large relative to skew.

The simulation presented in Fig. 4 illustrates the behavior of the tests when the data are skewed, as in the simulation in Section 3. The main difference from that simulation is that here the estimated rather than true variances are input to the publication bias tests. The performances of the two tests in this simulation are reversed from the results of the previous simulation. Egger's test is badly inconsistent, with the FPR approaching 1 by the time the sample size n is around 60. On the other hand, when the primary study sizes $m_i, i = 1, \ldots, n$, are $O(n^2)$, the FPR of the skew test is within about 2 percentage points of the nominal rate by a sample size of $n = 30$.

4.2 Studies based on binary data

The odds ratio, risk ratio, and risk difference are commonly used to report outcomes of studies based on binary data (Cooper et al., 2019). We focus on the odds ratio since a priori considerations as well as simulations have previously indicated that it poses difficulties for Egger's test (Macaskill et al., 2001). We also consider the risk ratio for the sake of comparison. We make the simplifying assumption that the primary studies are balanced.

For study i, $1 \leq i \leq n$, let $\overline{p}_{T,i} = \frac{1}{m_i}\sum_{j=1}^{m_i} T_{ij}E_{ij}$ and $\overline{p}_{C,i} = \frac{1}{m_i}\sum_{j=1}^{m_i}(1 - T_{ij})E_{ij}$ denote the observed proportions of the occurrences of an event in the treatment and control groups, each of size m_i, adopting the notation of (9). Let $p_{T,i}$ and $p_{C,i}$ denote the underlying class probabilities. We will also state results using an alternative parameterization,

$$\begin{aligned}\delta_i &= \operatorname{logit}(p_{T,i}) - \operatorname{logit}(p_{C,i}) \\ \mu_i &= (\operatorname{logit}(p_{T,i}) + \operatorname{logit}(p_{C,i}))/2.\end{aligned}$$

The first, which is the population log-odds in study i, measures the difference between $p_{T,i}$ and $p_{C,i}$, with 0 being no effect, and $\delta > 0$ or < 0 meaning treatment increases or reduces the chance of an event. The second parameter is a measure of the location of the average of the class probabilities, vanishing when $p_{T,i}$ and $p_{C,i}$ are equidistant from $1/2$, and > 0 or < 0 when their average is $>$ or $< 1/2$.

For both the risk ratio and odds ratio we find below that the overall story presented in the remarks following Proposition 3 hold. That is, under the proposition's conditions, the Egger bias, when non-vanishing, is $O(\sqrt{n})$, while the skew test bias is $\frac{1}{n}\sum_{i=1}^{n} O(\sqrt{n/m_i})$. However, the story needs to be qualified for notable situations when the bias does vanish and other times when the proposition assumptions do not hold and the bias diverges. We discuss the risk ratio first, where the behavior of the bias is more regular.

4.2.1 Risk ratio

The effect estimator for the risk ratio and its usual variance estimate under a balanced design are

$$\begin{aligned}Y_i &= \log(\overline{p}_{T,i}/\overline{p}_{C,i}) \\ \widehat{\operatorname{Var}}(Y_i) &= 1/R_i^2 = (1/\overline{p}_{T,i} + 1/\overline{p}_{C,i} - 2)/m_i.\end{aligned} \tag{13}$$

Analogous to Corollary 4, Corollary 5 approximates the publication bias test statistics for large m_i and n. We focus on the bias terms, leaving the asymptotic variances unevaluated. In the absence of a convenient form for the probability limit of the skew test statistic we further approximate it, for small μ and δ.

Corollary 5. *With (R_i, Y_i) as in (13), assume*

1. *the class probabilities $p_{C,i}$ and $p_{T,i}$, $i = 1, \ldots, n$, are uniformly bounded away from 0 and 1*
2. $\sup_n \operatorname{Corr}_n(Z, R) < 1$

Then,

$$\begin{aligned}T^{(E)}_{n,\{m_i\}}/\sqrt{n} = \operatorname{Var}(T^{(E)}_{n,\{m_i\}}/\sqrt{n})^{1/2}U_1 - \mathrm{P}_n \frac{\left(e^{\delta}-1\right)\left(e^{\frac{\delta}{2}+\mu}+e^{\delta}+1\right)}{2e^{\delta+\mu}+e^{\delta/2}\left(e^{\delta}+1\right)} \\ + O(\mathrm{E}_n(\mathrm{E}\,R^4 + \mathrm{E}\,Y^4 + \mathrm{E}\,Z^4)) + \mathrm{P}_n(O(1/\sqrt{m})) + O_P(1/n)\end{aligned}$$

and

$$T^{(s)}_{n,\{m_i\}}/\sqrt{n} = \operatorname{Var}(T^{(s)}_{n,\{m_i\}}/\sqrt{n})^{1/2}U_2 - \mathrm{P}_n\,\frac{1}{\sqrt{m}}\left((\frac{3}{\sqrt{2}}\mu + O(\mu^2))\delta + O(\delta^2)\right)$$
$$+ O(\mathrm{E}_n(\mathrm{E}\,R^6 + \mathrm{E}\,Y^6 + \mathrm{E}\,Z^6)) + \mathrm{P}_n(O(1/m)) + O_P(1/n).$$

As before, the bias of the skew but not Egger's test can be directly controlled by larger primary study sizes. In both cases the bias vanishes when there is no treatment effect, $\delta = 0$. As a result, the bias encountered in practice of even Egger's test may be small in balanced designs, when the treatment effect is on average close to the null, though the impact on FPR will unfortunately become worse as the sample size n increases; see Fig. 6. The bias increases in magnitude as $|\delta|$ increases from 0, approximately linearly initially then plateauing. See Fig. 5a plotting the surface of the bias, where $\delta = 0$ corresponds to the diagonal $p_C = p_T$ and is a line of anti-symmetry. Increased or decreased δ, corresponding to a harmful or protective effect, leads to a negative or positive bias, respectively, leading in turn, if the meta-analyst is performing a one-sided test, to an overly conservative or liberal test., and an overly liberal test for the more common two-sided test.

4.2.2 Odds ratio

The effect estimator for the odds ratio and its usual variance estimate under a balanced design are

$$\begin{aligned} Y_i &= \log\left(\left(\overline{p}_{T,i}(1-\overline{p}_{C,i})\right) / \left(\overline{p}_{C,i}\left(1-\overline{p}_{T,i}\right)\right)\right) \\ \widehat{\operatorname{Var}}(Y_i) = 1/R_i^2 &= (1/\overline{p}_{T,i} + 1/(1-\overline{p}_{T,i}) + 1/\overline{p}_{C,i} + 1/(1-\overline{p}_{C,i}))/m_i. \end{aligned} \tag{14}$$

Corollary 6. *With (R_i, Y_i) as in (14), assume*

1. *the class probabilities $p_{C,i}$ and $p_{T,i}$, $i = 1, \ldots, n$, are uniformly bounded away from 0 and 1*

2. $\sup_n \operatorname{Corr}_n(Z, R) < 1$

Then,

$$T^{(E)}_{n,\{m_i\}}/\sqrt{n} = \operatorname{Var}(T^{(E)}_{n,\{m_i\}}/\sqrt{n})^{1/2}U_1$$
$$- \mathrm{P}_n\,\frac{e^{\mu}\tanh\left(\frac{\delta}{2}\right)\left(\cosh\left(\frac{\delta}{2}\right)\cosh(2\mu) + \cosh(\mu)\right)\operatorname{sech}\left(\frac{1}{4}(\delta - 2\mu)\right)\operatorname{sech}\left(\frac{1}{4}(\delta + 2\mu)\right)}{|e^{2\mu} - 1|}$$
$$+ O(\mathrm{E}_n(\mathrm{E}\,R^4 + \mathrm{E}\,Y^4 + \mathrm{E}\,Z^4)) + \mathrm{P}_n(O(1/\sqrt{m})) + O_P(1/n).$$

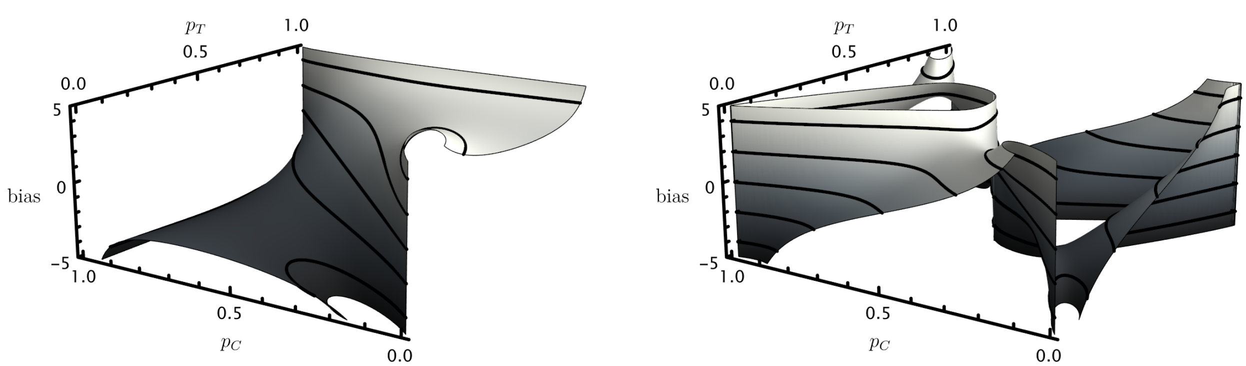
and

$$T^{(s)}_{n,\{m_i\}}/\sqrt{n} = \operatorname{Var}(T^{(s)}_{n,\{m_i\}}/\sqrt{n})^{1/2}U_2 - \mathrm{P}_n\,\frac{1}{\sqrt{m}}\left(\frac{3}{\sqrt{2}}\frac{\delta}{\operatorname{sign}(\mu)\mu^2} + O(\delta/|\mu|)\right)$$
$$+ O(\mathrm{E}_n(\mathrm{E}\,R^6 + \mathrm{E}\,Y^6 + \mathrm{E}\,Z^6)) + \mathrm{P}_n(O(1/m)) + O_P(1/n).$$

The behavior with the odds ratio is similar to the risk ratio in that the $\mathrm{P}_n\, O(1/\sqrt{m})$ factor can control the bias of the skew statistic. Further, $\delta = 0$ is a line of anti-symmetry, so, at least when $\mu \neq 0$, the biases vanish at $\delta = 0$. However, unlike the risk ratio, the bias diverges near $\mu = 0$, corresponding to class probabilities equidistant from 1/2, and can be arbitrarily large when the class probabilities p_T and p_C straddle 1/2. When $|\mu| > |\delta|/2$, corresponding to p_C and p_T lying on the same side of 1/2, the bias is well-behaved and plateaus. See Fig. 5 for a plot of the surface of the skew test bias in (p_T, p_C) space. The bias vanishes along the diagonal $\delta = 0, p_T = p_C$, blows up along anti-diagonal $\mu = 0, p_T + p_C = 1$, and pleateaus in the corners where $|\mu| > |\delta|/2$. Along the boundary between the blow-up along the anti-diagonal and the well-behaved plateaus are additional curves in (p_T, p_C) space, indicated in Fig. 5, where the bias vanishes.

Corollary 6 gives a quantitative description of the similar conclusion reached by simulation in Macaskill et al. (2001), namely, that there is no bias at the null while it increases away from the null. The corollary also reveals the blow-up effect near $\mu = 0$, which is obscured in the cited simulations since they randomly vary the class probabilities between monte carlo trials.

Fig. 6 presents the results of a simulation illustrating these points. Since the effect of the order of the primary sample sizes $m_1, \ldots, m_n$, is similar to the case of continuous data depicted in Fig. 4, we here set $m_i \sim O(n^2)$ and focus on the effect of μ. As expected in light of the foregoing remarks, the FPR of even the skew test can be arbitrarily poor for small enough μ, even when m_i are large relative to n.



(a) Risk ratio.

(b) Odds ratio.

Figure 5: Section 4.2 (Misspecified model with binary data): The unnormalized bias of the skew statistic based on binary data as a function of the underlying class probabilities. The classes are assumed balanced. The limit is taken for large n (number of studies) and m_i (the sizes of the primary studies). For both the RR and OR the bias vanishes along the diagonal, while the bias of the OR blows up along the anti-diagonal, corresponding to class probabilities straddling 1/2.

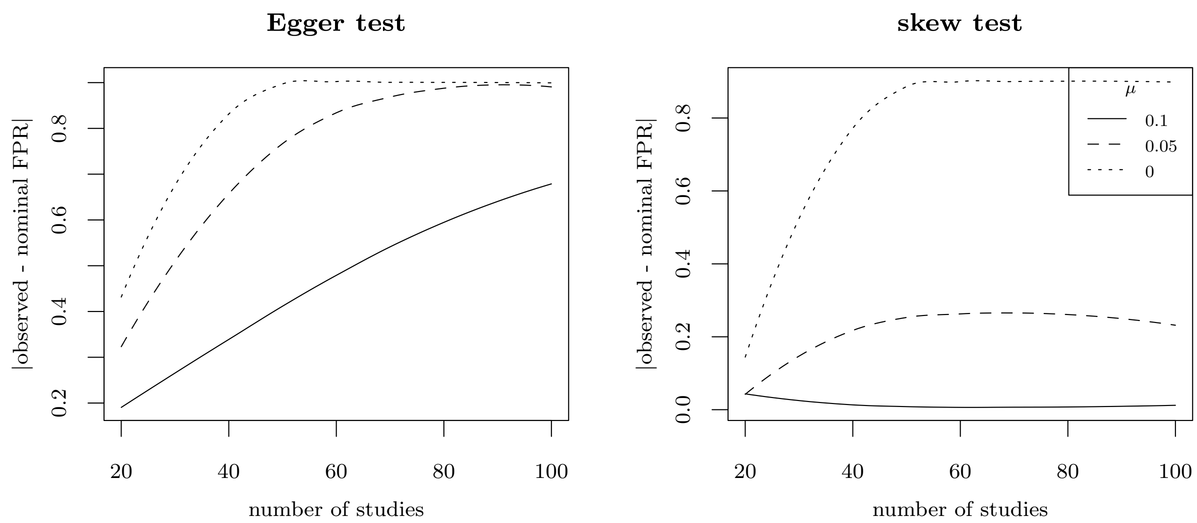


Figure 6: Section 4.2.2 (Misspecified model using the odds ratio): The effect of μ, measuring the location of the average class probabilities with respect to 1/2, on the FPR.

5 Discussion

There is a commonly used model of meta-analysis, which may include strong assumptions such as Gaussian effects and perfectly modeled study variances, to which Egger's test well-suited. Outside this model, e.g., where data have some asymmetry, the skew test appears more robust. Table 1 summarizes the results of the previous sections. It gives the order of the bias in the case that the studies are IID, i.e., ignoring the error terms due to heterogeneity in the study distributions in Proposition 3.

In several settings, such as the skew test under the mean model of Section 3 or Egger's test under the misspecified model, the bias is exacerbated by larger sample sizes. The conclusion, applicable to both asymptotic regimes, is that Egger's bias is $O(\sqrt{n})$, unless it vanishes completely, whereas the skew bias is $\frac{1}{n}\sum_{i=1}^{n} O(\sqrt{n/m_i})$. The conditions under which the bias of Egger's test vanishes generally involve symmetry in the data, or, more generally, uncorrelatedness of the study means and the reported reciprocal SDs. Though it may be unreasonable to expect any practically encountered data to be perfectly symmetric or skew-free, the results of Section 4 show that it is enough for the average skew to vanish quickly relative to the number of studies.

future work:

1. power 2. performance when combining the two tests

The software used to carry out the simulations and prepare the figures is publicly available at the author's website.

appendix:

other 3d figures of biases

sim showing effect of mi size on bias for count data–log-odds of RR.

Model	Egger's Test Bias	Skew Test Bias
n asymptotics		
Gaussian model	$O(1/n)$	$O(1/n)$
Symmetric distribution, true variances	$O(1/n)$	$O(1/n)$
Mean model	$O(1/\sqrt{n})$	$O(\sqrt{n})$
$n, \{m_i\}$ asymptotics		
SMDs, skew-free distribution	$O(1/n)$	$O(1/n)$
SMDs	$O(\sqrt{n})$	$\mathrm{P}_n \sqrt{n/m}$
OR, RR, no treatment effect	$O(1/n)$	$O(1/n)$
OR, RR	$O(\sqrt{n})$	$\mathrm{P}_n \sqrt{n/m}$

Table 1: Bias of Egger's test and the skew test.

References

Begg, C. and M. Mazumdar (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 1088–1101.

Bhattacharya, R. N., J. K. Ghosh, et al. (1978). On the validity of the formal edgeworth expansion. *Annals of Statist 6*(2), 434–451.

Cooper, H., L. V. Hedges, and J. C. Valentine (2019). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation.

Egger, M., G. D. Smith, M. Schneider, and C. Minder (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ 315*(7109), 629–634.

Giné, E., F. Götze, and D. M. Mason (1997). When is the student t-statistic asymptotically standard normal? *The Annals of Probability 25*(3), 1514–1531.

Higgins, J. P., I. R. White, and J. Anzures-Cabrera (2008). Meta-analysis of skewed data: combining results reported on log-transformed or raw scales. *Statistics in medicine 27*(29), 6072–6092.

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

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

Michael, H. (2025). The effect of screening for publication bias on the outcomes of meta-analyses. *Scandinavian Journal of Statistics 52*(1), 513–531.

Rothstein, H. R., A. J. Sutton, and M. Borenstein (2005). *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. West Sussex, England: John Wiley & Sons.

Takemura, A. and K. Takeuchi (1988). Some results on univariate and multivariate cornish-fisher expansion: algebraic properties and validity. *Sankhyā: The Indian Journal of Statistics, Series A 50*, 111–136.

Wilkins, J. E. (1944). A note on skewness and kurtosis. *The Annals of Mathematical Statistics 15*(3), 333–335.

Proof of Proposition 1. Under the gaussian model (5) the Egger test statistic, being the t-statistic for the intercept in the regression (3), follows Student's t distribution with $n-2$ degrees of freedom. The expansion given in the proposition is the Edgeworth expansion for this distribution.

The formal Edgeworth expansion for $T_n^{(s)}$, with two correction terms, is given by

$$\mathrm{P}(T_n^{(s)} < x) = \Phi(x) - \phi(x)\left(\frac{1}{2n}ax + \frac{1}{6}\kappa_3 H_2(x) + \frac{1}{24}\kappa_4 H_3(x) + \frac{1}{72}\kappa_3^2 H_5(x)\right) + o(1/n), \tag{15}$$

where κ_j are the cumulants of $T_n^{(s)}$, a is given by the $1/n$ term in the variance approximation $\mathrm{Var}(T_n^{(s)}) = 1 + a/n + o(1/n)$, and H_j is the j^{th} probabilist's Hermite polynomial. To approximate the first four cumulants with error $o(1/n)$ we expand the expression for $T_n^{(s)}$ given in (4), viewed as a function of $(\hat{\theta}_2, \hat{\theta}_3)$, in a Taylor series about about the means $(\theta_2, \theta_3) = \mathrm{E}(\hat{\theta}_2, \hat{\theta}_3) = (1 - 2/n, 0)$. We then substitute the moments of the Taylor approximation, computed to order $1/n$. The validity of the approximation (15) and this procedure are established for smooth functions of a mean by Bhattacharya et al. (1978). The proof of Proposition 3 below shows how $T_n^{(s)}$ may be expressed as a smooth function of means.

The odd moments of $T_n^{(s)}$ vanish since the residuals are symmetric under (5). Let $\hat{Z}_2 = \sqrt{n}(\hat{\theta}_2 - (1 - n/2))$ and $\hat{Z}_3 = \sqrt{n}\hat{\theta}_3$. With this notation,

$$T_n^{(s)} = \frac{1}{\sqrt{6}}\hat{Z}_3 - \frac{\sqrt{3}}{2\sqrt{2n}}\hat{Z}_2\hat{Z}_3 + \frac{\sqrt{3}}{8\sqrt{2n}}(4 + 5\hat{Z}_2^2)\hat{Z}_3 + o_P(1/n).$$

The second and fourth moments are given by

$$\begin{aligned}
\mathrm{E}(T_n^{(s)})^2 &= \frac{1}{6}\,\mathrm{E}\,\hat{Z}_3^2 - \frac{1}{2\sqrt{n}}\,\mathrm{E}(\hat{Z}_2\hat{Z}_3^2) + \frac{1}{n}\left(\frac{1}{2}\,\mathrm{E}\,\hat{Z}_3^2 + \mathrm{E}(\hat{Z}_2^2\hat{Z}_3^2)\right) + o(1/n) \\
\mathrm{E}(T_n^{(s)})^4 &= \frac{1}{36}\,\mathrm{E}\,\hat{Z}_3^4 - \frac{1}{6\sqrt{n}}\,\mathrm{E}(\hat{Z}_2\hat{Z}_3^4) + \frac{1}{n}\left(\frac{1}{6}\,\mathrm{E}\,\hat{Z}_3^4 + \frac{7}{12}\,\mathrm{E}(\hat{Z}_2^2\hat{Z}_3^4)\right) + o(1/n).
\end{aligned} \tag{16}$$

We compute the mixed moments of $\hat{Z}_2$ and $\hat{Z}_3$ appearing in (16) directly. The regressand $Y/\hat{\sigma}$ in (3) is distributed as $N(\theta/\hat{\sigma}, 1)$ given the regressors $1/\hat{\sigma}$. A location shift $Y_1, \ldots, Y_n$, only adds to the regressand a vector lying in the span of the design. It follows that θ can be taken to be 0 without affecting the test statistics. With $\theta = 0$ the regressand is distributed as standard normal given the regressors and therefore independent of the regressors. In sum, denoting $Z_i = Y_i/\hat{\sigma}_i$, the regression (3) under the Gaussian model (5) can be written

$$\begin{aligned}
&Z_i \text{ vs. intercept and } R_i, i = 1, \ldots, n, \\
&Z \sim N(0, 1), \\
&Z \perp\!\!\!\perp R.
\end{aligned}$$

In this form,

$$\hat{\epsilon}_i = Z_i - \overline{Z} - \hat{\beta}_1(R_i - \overline{R})$$
$$\hat{Z}_p = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Z_i - \overline{Z} - \hat{\beta}_1(R_i - \overline{R}))^p, p = 2, 3.$$

$\mathrm{E}\,\hat{Z}_3^2$ and $\mathrm{E}\,\hat{Z}_3^4$ are $6 - 36/n$ and $108(1+18/n)$ respectively. For $\mathrm{E}(\hat{Z}_2\hat{Z}_3^2)$ and $\mathrm{E}(\hat{Z}_2\hat{Z}_3^4)$, (16) only requires approximations up to $o(1/\sqrt{n})$ error, which are given by $36/\sqrt{n}$ and $1296/\sqrt{n}$. Only the constant terms are needed for $\mathrm{E}(\hat{Z}_2^2\hat{Z}_3^2)$ and $\mathrm{E}(\hat{Z}_2^2\hat{Z}_3^4)$, which can be obtained as the probability limits 12 and 216. Substituting these values in (16), $\mathrm{E}(T_n^{(s)})^2 = 1 - 9/n + o(1/n)$ and $\mathrm{E}(T_n^{(s)})^4 = 3 - 34/n + o(1/n)$. Therefore the cumulants are given by $\kappa_1 = \kappa_3 = 0$, $\kappa_2 = 1 - 9/n$, and $\kappa_4 = 20/n$. Substituting these values for κ_j into (15) returns the expansion for $T_n^{(s)}$ given in the proposition.

□

Proof of Proposition 2. The i^{th} residual, $1 \le i \le n$, in the Egger regression (7) is

$$\begin{aligned}\hat{\epsilon}_i &= Y_i - \left(\overline{Y}(\overline{1/\hat{\sigma}^2} - \overline{1/\hat{\sigma}}1/\hat{\sigma}_i) + \overline{Y/\hat{\sigma}}\left(1/\hat{\sigma}_i - \overline{1/\hat{\sigma}}\right)\right)/\hat{V}(1/\hat{\sigma}) \\ &= 1/\hat{\sigma}_i(Y_i + c_1) + c_2,\end{aligned}$$

where $c_1 = (\overline{1/\hat{\sigma}}\;\overline{Y/\hat{\sigma}} - \overline{Y/\hat{\sigma}^2})/\hat{V}(1/\hat{\sigma})$, $c_2 = (\overline{1/\hat{\sigma}}\;\overline{Y/\hat{\sigma}^2} - \overline{1/\hat{\sigma}^2}\;\overline{Y/\hat{\sigma}})/\hat{V}(1/\hat{\sigma})$, and $\hat{V}(1/\hat{\sigma}) = \overline{1/\hat{\sigma}^2} - \overline{1/\hat{\sigma}}^2$. By the moment model assumptions (6),

$$\hat{V}(1/\hat{\sigma})c_1 = \mathrm{E}(1/\hat{\sigma})\,\mathrm{E}(Y/\hat{\sigma}) - \mathrm{E}(Y/\hat{\sigma}^2) + O_P(1/\sqrt{n}) = -\theta(\mathrm{E}(1/\hat{\sigma}^2) - \mathrm{E}(1/\hat{\sigma})^2) + O_P(1/\sqrt{n})$$

and

$$\hat{V}(1/\hat{\sigma})c_2 = \mathrm{E}(1/\hat{\sigma})\,\mathrm{E}(Y/\hat{\sigma}^2) - \mathrm{E}(1/\hat{\sigma}^2)\,\mathrm{E}(Y/\hat{\sigma}) + O_P(1/\sqrt{n}) = O_P(1/\sqrt{n}).$$

The assumption that $\hat{V}(1/\hat{\sigma})$ has a probability limit > 0 then implies $c_1 = -\theta + O_P(1/\sqrt{n})$ and $c_2 = O_P(1/\sqrt{n})$.

The p^{th} sample moment of the residuals is

$$\frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i^p = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=0}^{p}\binom{p}{j} 1/\hat{\sigma}_i^j (Y_i + c_1)^j c_2^{p-j}. \tag{17}$$

Since $c_2 = O_P(1/\sqrt{n})$ the terms in the inner sum of (17) with $j \le p-2$ are $o_P(1/\sqrt{n})$ as long as $\sum_i 1/\hat{\sigma}_i^j$ and $\sum_i (Y_i/\hat{\sigma}_i)^j$ are $O_P(1)$, which has been assumed for $p \le 3$. Therefore,

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i^p &= o_P(1/\sqrt{n}) + pc_2 \frac{1}{n}\sum_i 1/\hat{\sigma}_i^{p-1}(Y_i + c_1)^{p-1} + \frac{1}{n}\sum_i 1/\hat{\sigma}_i^p (Y_i + c_1)^p \\
&= o_P(1/\sqrt{n}) + pc_2 \frac{1}{n}\sum_i 1/\hat{\sigma}_i^{p-1}((Y_i - \theta)^{p-1} + (p-1)(Y_i - \theta)^{p-2}(c_1 + \theta)) \\
&+ \frac{1}{n}\sum_i 1/\hat{\sigma}_i^p((Y_i - \theta)^p + p(Y_i - \theta)^{p-1}(c_1 + \theta)) \\
&= o_P(1/\sqrt{n}) + \frac{1}{n}\sum_i (1/\hat{\sigma}_i(Y_i - \theta))^p + pc_2 \frac{1}{n}\sum_i \mathrm{E}(1/\hat{\sigma}_i(Y_i - \theta))^{p-1} \\
&+ p(c_1 + \theta)\frac{1}{n}\sum_i \mathrm{E}(1/\hat{\sigma}_i^p(Y_i - \theta)^{p-1}).
\end{aligned}
\tag{18}
$$

Applying again the moment assumptions (6), the last two terms in (18) are

$$
\begin{aligned}
&\frac{pc_2}{n}\sum_i \mathrm{E}(1/\hat{\sigma}_i(Y_i - \theta))^{p-1} + \frac{p(c_1 + \theta)}{n}\sum_i \mathrm{E}(1/\hat{\sigma}_i^p(Y_i - \theta)^{p-1}) \\
&\qquad = \begin{cases} 0 & \text{for } p = 2 \\ 3(\theta\overline{1/\hat{\sigma}} - \overline{Y/\hat{\sigma}}) + o_P(1/\sqrt{n}) & \text{for } p = 3. \end{cases}
\end{aligned}
$$

Substituting the linearizations of the $p = 2$ and $p = 3$ sample moments computed above into a Taylor expansion about the probability limits (θ_2, θ_3),

$$
\frac{\frac{1}{n}\sum_i \hat{\epsilon}_i^3}{(\frac{1}{n}\sum_i \hat{\epsilon}_i^2)^{3/2}} = o_P(1/\sqrt{n}) + \theta_2^{-3/2} \cdot (\text{linearization at p=3}) - \frac{3}{2}\theta_2^{-5/2}\theta_3 \cdot ((\text{linearization at p=2}) - \theta_2),
$$

gives the expansion stated in the proposition. □

Proof of Proposition 3. For the purposes of the proof, each of the two test statistics is a smooth function of averages of smooth functions of averages of the data, the first average being in n and the second being in the m_i; the first smooth function being the test statistic itself, and the second set of smooth functions being polynomials composed with the Y_i and R_i.

Let n independent pairs $(R_i, Y_i), i = 1, \ldots, n$ be given. Let $h^{(uv)}(R_i, Y_i) = R_i^u Y_i^v, 0 \le u \le u_0, 0 \le v \le v_0, 1 \le i \le n$, for integers u and v, and let $h(R_i, Y_i) = (h^{(0,0)}(R_i, Y_i), \ldots, h^{(u_0,v_0)}(R_i, Y_i))$ collect the $h^{(u,v)}$ in a vector of length $(u_0 + 1)(v_0 + 1)$. Let $\overline{h}_n = \frac{1}{n}\sum_{i=1}^n h(R_i, Y_i)$ denote the vector of sample mixed moments of (R_i, Y_i). Let g be a smooth scalar function of $\overline{h}_n$.

Lemma 7. *With the above data generating process, assume:*

1. *For each i, (R_i, Y_i) is a smooth function of the average of m_i independent random variables*

2. $\sup_i \mathrm{E}\,|h^{(u,v)}(R_i, Y_i)|^{2+\delta} < \infty$ *as* $n \to \infty$*, for* $0 \le u \le u_0, 0 \le v \le v_0$*, with* u_0 *and* v_0 *both* ≥ 2

3. $g''(\mathrm{P}_n \,\mathrm{E}\, h(R, Y)) = O(1)$.

Then

$$\begin{aligned} g(\overline{h}_n) &= (g'(\mathrm{E}\,\overline{h}_n)^T(\mathrm{Cov}\,\overline{h}_n)g'(\mathrm{E}\,\overline{h}_n))^{1/2}U + \mathrm{P}_n\, g(\mathrm{E}\, h(R, Y)) \\ &\quad + O(\mathrm{E}_n(\mathrm{E}\, R^{u_0+v_0} + \mathrm{E}\, Y^{u_0+v_0} + \mathrm{E}\, Z^{u_0+v_0})) + O_P(1/n). \end{aligned}$$

Proof of Lemma 7. Decompose

$$g(\overline{h}_n) = g(\overline{h}_n) - g(\mathrm{E}\,\overline{h}_n) + g(\mathrm{E}\,\overline{h}_n). \tag{19}$$

Assumptions 1 and 2 imply the Lyapunov condition

$$\frac{1}{n^{1+\delta/2}} \sum_{i=1}^{n} \mathrm{E}\,|h_i - \mathrm{E}\, h_i|^{2+\delta} \to 0.$$

Therefore, there is a standard multivariate Gaussian U such that

$$\sqrt{n}(\overline{h}_n - \mathrm{E}\,\overline{h}_n) = \mathrm{Cov}(\sqrt{n}\overline{h}_n)^{1/2}U + O_P(1/\sqrt{n}).$$

By the delta method,

$$g(\overline{h}_n) = g(\mathrm{E}\,\overline{h}_n) + (g'(\mathrm{E}\,\overline{h}_n)^T(\mathrm{Cov}\,\overline{h}_n)g'(\mathrm{E}\,\overline{h}_n))^{1/2}U + O_P(1/n).$$

For the non-random last term in (19), i.e.,

$$g(\mathrm{E}\,\overline{h}_n) = g(\mathrm{E}(\frac{1}{n}\sum_{i=1}^{n} h(R_i, Y_i))) = g(\mathrm{P}_n\,\mathrm{E}\, h(R, Y)),$$

we interchange the g and P_n, putting the difference in an error term. Expanding with i fixed,

$$\begin{aligned} g(\mathrm{E}\, h(R_i, Y_i)) &= g(\mathrm{P}_n\,\mathrm{E}\, h(R, Y)) + (\mathrm{E}\, h(R_i, Y_i) - \mathrm{P}_n\,\mathrm{E}\, h(R, Y))g'(\mathrm{P}_n\,\mathrm{E}\, h(R, Y)) \\ &\quad + \frac{1}{2}(\mathrm{E}\, h(R_i, Y_i) - \mathrm{P}_n\,\mathrm{E}\, h(R, Y))^T g''(\mathrm{P}_n\,\mathrm{E}\, h(R, Y))(\mathrm{E}\, h(R_i, Y_i) - \mathrm{P}_n\,\mathrm{E}\, h(R, Y)) \\ &\quad + o(|\,\mathrm{E}\, h(R_i, Y_i) - \mathrm{P}_n\,\mathrm{E}\, h(R, Y)|^2). \end{aligned}$$

Taking the empirical average of each side yields

$$\mathrm{P}_n\, g(\mathrm{E}\, h(R, Y))) - g(\mathrm{P}_n\,\mathrm{E}\, h(R, Y)) = O(\mathrm{Var}_n\,\mathrm{E}\, h(R, Y)),$$

provided $g''(\mathrm{P}_n\,\mathrm{E}\, h(R, Y))$ is $O(1)$, as has been assumed. □

Next we apply the lemma to the Egger test statistic $T^{(E)}_{n,\{m_i\}}$. To express the test statistic as a smooth function of means, let $h = (h^{(01)}, \ldots, h^{(22)}) : \mathbb{R}^2 \to \mathbb{R}^5$, where for $1 \leq j \leq 2, 0 \leq k \leq j$, $h^{(jk)} : \mathbb{R}^2 \to \mathbb{R}$ is given by $(x_1, x_2) \mapsto x_1^j x_2^k$. Let $g : \mathbb{R}^5 \to \mathbb{R}$ be given by

$$g : (x_{01}, \ldots, x_{22}) \mapsto \frac{x_{11}x_{20} - x_{10}x_{21}}{\sqrt{x_{20}\left(2x_{10}x_{21}x_{11} - x_{20}x_{11}^2 - x_{21}^2 - \left(x_{10}^2 - x_{20}\right)x_{22}\right)}}.$$

Then

$$T^{(E)}_{n,\{m_i\}}/\sqrt{n} = g \circ \frac{1}{n}\sum_{i=1}^{n} h(Y_i, R_i). \tag{20}$$

Assumptions 1 and 2 of the Lemma are assumed directly by Proposition 3. Assumption 3 of the Lemma requires the hessian of the unnormalized Egger statistic g, evaluated at the empirical averages of the moments of $(R_i, Y_i), i = 1, \ldots, n$, be bounded as $n \to \infty$. Write $T^{(E)}_{n,\{m_i\}}/\sqrt{n}$ as $N/\sqrt{D}$ with N and D in the components of $\overline{h}_n$,

$$g'' = N''D^{-1/2} - N'D'D^{-3/2} - \frac{1}{2}ND''D^{-3/2} + \frac{3}{4}N(D')^2D^{-5/2}.$$

Due to Assumption 2 all terms are bounded other than the negative powers of D, so it suffices to ensure that D is bounded away from 0. This condition follows since D represents the variance of the intercept t-statistic,

$$SSE \cdot \frac{\mathrm{P}_n R^2}{\mathrm{P}_n R^2 - (\mathrm{P}_n R)^2},$$

Assumption 3 implies the second factor is bounded away from 0 and assumption 4 implies that the sum of squared errors of (3) is bounded away from 0.

To simplify the bias term $g(\mathrm{E}\, h(R, Y)$ appearing in 7, we approximate the mixed moments appearing in (20) of order > 2 to order $o(1/m_i^2)$ using moments of order ≤ 2. These higher order moments are $\mathrm{E}\, R_i^2 Y_i$ and $\mathrm{E}\, R_i^2 Y_i^2$, $i = 1, \ldots, n$. Since it is assumed that S and Y are smooth functions of means,

$$\mathrm{E}\, R_i^2 Y_i = (\mathrm{E}\, R_i)^2\, \mathrm{E}\, Y_i + \frac{1}{m}(2\,\mathrm{E}\, R_i \operatorname{Cov}(R_i, Y_i) + \mathrm{E}\, Y_i \operatorname{Var} R_i) + \frac{1}{m^2}\,\mathrm{E}((R_i - \mathrm{E}\, R_i)^2(Y_i - \mathrm{E}\, Y_i)) + o(\frac{1}{m^2})$$

$$\mathrm{E}\, R_i^2 Y_i^2 = (\mathrm{E}\, R_i)^2(\mathrm{E}\, Y_i)^2 + \frac{1}{m}((\mathrm{E}\, R_i)^2 \operatorname{Var} Y_i + (\mathrm{E}\, Y_i)^2 \operatorname{Var} R_i + \mathrm{E}\, R_i\, \mathrm{E}\, Y_i \operatorname{Cov}(R_i, Y_i))$$

$$+ \frac{1}{m^2}(\operatorname{Var} R_i \operatorname{Var} Y_i + 2\operatorname{Cov}(R_i, Y_i)^2 + 2\,\mathrm{E}\, R_i\, \mathrm{E}((R_i - \mathrm{E}\, R_i)(Y_i - \mathrm{E}\, Y_i)^2) + 2\,\mathrm{E}\, Y_i\, \mathrm{E}((R_i - \mathrm{E}\, R_i)^2(Y_i - \mathrm{E}\, Y_i)$$

$$+ o(\frac{1}{m^2}).$$

Substituting these approximations into $g(\mathrm{E}\, h(R, Y)$ yields for the numerator

$$-\frac{1}{m_i}(\mathrm{E}\, R_i)^2 \operatorname{Cov}(R_i, Y_i) + o_P(\frac{1}{m_i}),$$

for the denominator,

$$\frac{1}{m_i}(\mathrm{E}\,R_i)^2(\mathrm{Var}\,R_i\,\mathrm{Var}\,Y_i - \mathrm{Cov}(R_i,Y_i))^{1/2} + o_P(\frac{1}{m_i}),$$

and for their ratio,

$$-\frac{\mathrm{Corr}(R_i,Y_i)}{\sqrt{1-\mathrm{Corr}(R_i,Y_i)}} + o_P(\frac{1}{m_i}).$$

Finally, we apply the Lemma to the skew statistic, following the same steps. To express the test statistic as a smooth function of means, let $h = (h^{(01)}, \ldots, h^{(33)}) : \mathbb{R}^2 \to \mathbb{R}^{10}$, where for $0 \le j \le 3, 0 \le k \le j$, $h^{(jk)} : \mathbb{R}^2 \to \mathbb{R}$ is given as before by $(x_1, x_2) \mapsto x_1^j x_2^k$. Let $g : \mathbb{R}^{10} \to \mathbb{R}$ be given by

$$g : (x_{00}, \ldots, x_{33}) \mapsto \frac{m_3(x_{01}, \ldots, x_{33})}{\sqrt{6}m_2(x_{01}, \ldots, x_{33})^{3/2}}$$

where for $p = 2, 3$, $m_p : \mathbb{R}^{(p+1)(p+2)/2} \to \mathbb{R}$, representing the p^{th} sample moment of the residuals, is given by

$$\begin{aligned}(x_{00}, \ldots, x_{pp}) \mapsto \sum_{j=0}^{p} \binom{p}{j} c_2^{p-j} \sum_{k=0}^{j} \binom{j}{k} c_1^{j-k} x_{jk},\\ \text{with } c_1 = (x_{10}x_{11} - x_{21})/(x_{20} - x_{10}^2),\\ c_2 = (x_{10}x_{21} - x_{20}x_{11})/(x_{20} - x_{10}^2).\end{aligned}$$

Then

$$T^{(s)}_{n,\{m_i\}} = g \circ \frac{1}{n}\sum_{i=1}^{n} h(Y_i, R_i). \tag{21}$$

Assumption 3 of the lemma follows as with the Egger statistic from Assumption 4 of the Proposition, since the denominator of the skew statistic is a power of the sum of squared errors of the Egger regression.

To simplify the bias term $g(\mathrm{E}\,h(R,Y)$ appearing in 7, rather than approximating the mixed moments $\mathrm{E}\,R_i^j Y_i^k$ directly as we did with the Egger statistic above, we approximate (R_i, Y_i) by an Edgeworth-type stochastic expansion. Specifically, let

$$(Z_{1i}, Z_{2i}) = \mathrm{Cov}((R_i, Y_i))^{1/2}(R_i - \mathrm{E}\,R_i, Y_i - \mathrm{E}\,Y_i)$$

center and standardize (R_i, Y_i), where $\mathrm{Cov}((R_i, Y_i))^{1/2}$ is the Cholesky decomposition of the covariance matrix of (R_i, Y_i). Then Z_{2i} is distributed as the standardized residual ϵ_i defined in Proposition 3. By Assumptions 1 and 2, (Z_{1i}, Z_{2i}) admits an expansion in powers of $1/\sqrt{m_i}$ and coefficients that are polynomials in IID standard normals U_{i1}, U_{i2} (Bhattacharya et al., 1978; Takemura and Takeuchi, 1988). Direct substitution of this approximation in (21) returns the bias term $\mathrm{E}\,\epsilon_i^3/\sqrt{6}$ appearing in (10).

□

Proof of Corollary 4. The assumptions of Proposition 3 to be verified are 2 and 3. Below we first show that Assumptions 5 and 5 of the corollary imply the moments of R_i are bounded uniformly over i for large enough m_i. From there along with Assumption 2 of the corollary, we show that the moments of $Z_i = Y_i R_i$ are bounded uniformly. We then show that the mixed moments of (R_i, Y_i) are bounded uniformly. Finally we show that the empirical second moment of R_i is bounded away from 0.

Unable to find a handy reference We argue as follows that the boundedness assumption imposed on the densities of the observations implies the reciprocal sample SD based on a sample $W_{i1}, \ldots, W_{im_i}$, has a k^{th} moment when $m > k+1 \geq 2$. We omit the subscript i in this argument. Since

$$
\begin{aligned}
\mathrm{E}((\overline{W^2} - \overline{W}^2)^{-k/2}) &= \frac{k}{2}\int_0^\infty u^{-k/2-1}\,\mathrm{P}(\overline{W^2} - \overline{W}^2 < u)du \\
&\leq \epsilon^{-k/2} + \frac{k}{2}\int_0^\epsilon u^{-k/2-1}\,\mathrm{P}(\overline{W^2} - \overline{W}^2 < u)du \qquad (22)
\end{aligned}
$$

the k^{th} moment of the reciprocal sample SD can be controlled by bounding the CDF of the sample variance near the origin. maybe cite old agresti paper Regarding the sample variance as the scaled L_2 norm of the projection of the data vector $(W_1, \ldots, W_m) \in \mathbb{R}^m$ onto the orthogonal complement of the vector of ones $\mathbb{1} = (1, 1, \ldots, 1)$, that CDF is

$$
\mathrm{P}(\overline{W^2} - \overline{W}^2 < u) = \mathrm{P}(\{x + v : x \propto \mathbb{1}, \sum_{i=1}^m v_i = 0, |v|_2 \leq \sqrt{mu}\}).
$$

The RHS is the mass in the cylinder with axis $\mathbb{1}$ and base an $m-1$-dimensional ball of radius $\sqrt{mu}$. To compute this probability we integrate the joint density on the cylinder's sections by hyperplanes through the x_m axis. The permutation symmetry of the coordinates in this cylinder implies that the region obtained by fixing a value for x_m is an $m-1$-dimensional ball of radius $r = r(m, u) = \sqrt{m/(m-1)}\sqrt{mu}$,

$$
\begin{aligned}
&\mathrm{P}(\{x + v : x \propto \mathbb{1}, \sum_{i=1}^m v_i = 0, |v|_2 \leq \sqrt{mu}\}) \\
&= \int_{-\infty}^\infty dx_m f(x_m) \int_{x_m \mathbb{1} + rB_{m-1}} dx_1 \cdots dx_{m-1} f(x_1) \cdots f(x_{m-1}).
\end{aligned}
$$

In the region of integration of the inner integral above, B_{m-1} denotes the unit ball in $\mathbb{R}^{m-1}$ embedded in the first $m-1$ coordinates of $\mathbb{R}^m$. Since the integrand of the inner integral doesn't depend on x_m we can take this region of integration to be $(x_m, \ldots, x_m, 0) + rB_{m-1}$. Consider the function

$$
x \mapsto \int_{(x,\ldots,x,0)+rB_{m-1}} dx_1 \cdots dx_{m-1} f(x_1) \cdots f(x_{m-1})
$$

for real x in the support of f. Since f is assumed continuous on a closed interval containing this union and, in case this interval is unbounded, are assumed to tend to 0 as $|x| \to \infty$, the defined function is also continuous on the interval and tends to 0 as $|x| \to \infty$, and so achieves a maximum at some $x^* \in \mathbb{R}$. Therefore,

$$
\begin{aligned}
&\int_{-\infty}^{\infty} dx_m f(x_m) \int_{x_m \mathbb{1} + rB_{m-1}} dx_1 \cdots dx_{m-1} f(x_1) \cdots f(x_{m-1}) \\
&= \int_{-\infty}^{\infty} dx_m f(x_m) \int_{(x_m,\ldots,x_m,0)+rB_{m-1}} dx_1 \cdots dx_{m-1} f(x_1) \cdots f(x_{m-1}) \\
&\leq \int_{-\infty}^{\infty} dx_m f(x_m) \int_{(x^*,\ldots,x^*,0)+rB_{m-1}} dx_1 \cdots dx_{m-1} f(x_1) \cdots f(x_{m-1}) \\
&= \int_{(x^*,\ldots,x^*,0)+rB_{m-1}} dx_1 \cdots dx_{m-1} f(x_1) \cdots f(x_{m-1}) \\
&\leq (|f|_\infty)^m \mathrm{Vol}(B_{m-1}) r^{m-1} \\
&= (|f|_\infty)^m \mathrm{Vol}(B_{m-1}) (m^2 u/(m-1))^{((m-1)/2)}.
\end{aligned}
$$

Above, $\mathrm{Vol}(B_{m-1}) = \pi^{n/2}/\Gamma(n/2+1)$ refers to the volume of the unit ball in $\mathbb{R}^{m-1}$. Applying Stirling's approximation to this expression,

$$\mathrm{P}(\overline{W^2} - \overline{W}^2 < u) \leq c\,(|f|_\infty)^m \frac{(2\pi \mathrm{e})^{m/2}}{\sqrt{m}} u^{(m-1)/2},$$

where c is a constant that doesn't depend on m. Substituting this CDF bound in (22), and supposing $m > k+1$,

$$\mathrm{E}\, R^k \leq \epsilon^{-k/2} + c\,\frac{k}{m-k-1} \epsilon^{-1-k/2} (|f|_\infty \sqrt{2\pi \mathrm{e}} \sqrt{\epsilon})^m.$$

Taking $\epsilon < 1/(2\pi \mathrm{e} |f|_\infty^2)$ ensures this bound doesn't grow with m and also shows that $\sup_i \mathrm{E}\, R_i^k$ is finite when $\sup_i |f_i|_\infty$ is bounded, as has been assumed.

Next, we establish Assumption 2 of the Proposition when $j = k$, i.e., uniform bounds on the simple moments of $Z = RY$. When $j = k = 2$, by the Cauchy-Schwarz inequality,

$$\mathrm{E}\, Z_i^2 = \mathrm{E}\, R_i^2 Y_i^2 \leq (\mathrm{E}\, R_i^4)^{1/2} (\mathrm{E}\, Y_i^4)^{1/2}.$$

It was established above that under the Corollary's conditions $\sup_i \mathrm{E}\, R_i^4 < \infty$, while by Assumption 2,

$$\mathrm{E}\, Y_i^4 = \frac{1}{m_i^4} \mathrm{E}(\sum_{j=1}^{m_i} W_{ij})^4 \leq \sup_{i,j} \mathrm{E}\, W_{ij}^4 < \infty.$$

It then follows by Giné et al. (1997, Theorem 2.5) that $Z_j = Y_j S_j$ have uniformly bounded positive moments.

For Assumption 2 of the Proposition when $j \neq k$, i.e., the mixed moments of R_i and Y_i, by Hölder's inequality,

$$\mathrm{E}((Y^j R^k)^{2+\delta}) = \mathrm{E}((Z^j R^{k-j})^{2+\delta}) \leq |Z^{(2+\delta)j}|_p |R^{(2+\delta)(k-j)}|_q,$$

where $1/p + 1/q = 1, p > 1, q > 1$. Since by the last part we can take any $p > 1$, a uniform bound on the mixed moment follows from a uniform finite bound on $\mathrm{E}\, R^{(2+\delta)(k-j)}$ for some $\delta > 0$, which has also been established above.

Finally, for Assumption 3 of the Proposition, by Jensen's inequality,

$$\mathrm{P}_n R^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\overline{W_{i\cdot}^2}-(\overline{W_{i\cdot}})^2} \geq (\frac{1}{n}\sum_{i=1}^{n}\overline{W_{i\cdot}^2}-\frac{1}{n}\sum_{i=1}^{n}(\overline{W_{i\cdot}})^2)^{-1} \geq (\frac{1}{n}\sum_{i=1}^{n}\overline{W_{i\cdot}^2})^{-1}.$$

Since $\frac{1}{n}\sum_{i=1}^{n}\overline{W_{i\cdot}^2} \to \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\,\overline{W_{i\cdot}^2} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i}\mathrm{E}\,W_{ij}^2$, the uniform bound on $\mathrm{E}\,W_{ij}^2$ given by Assumption 2 of the corollary implies the empirical second moments of R_i are bounded.

With the conditions of Proposition 3 verified, the formulas for the asymptotic variance and bias can be computed by substituting into the general formula given by the Proposition Taylor expansions for R_i and Y_i,

$$\begin{aligned}
R_i &= \frac{1}{\sqrt{\mathrm{E}\,W_i^2-(\mathrm{E}\,W_i)^2}} + \frac{2d_{1,i}\,\mathrm{E}\,W_i - d_{2,i}}{2(\mathrm{E}\,W_i^2-(\mathrm{E}\,W_i)^2)^{3/2}} \\
&\quad + \frac{4d_{1,i}^2(2(\mathrm{E}\,W_i)^2+\mathrm{E}\,W_i^2) - 12d_{1,i}d_{2,i}\,\mathrm{E}\,W_i + 3d_{2,i}^2}{8(\mathrm{E}\,W_i^2-(\mathrm{E}\,W_i)^2)^{5/2}} + O(d_{1,i}^3+d_{2,i}^3) \\
Y_i &= \mathrm{E}\,W_i + d_{1,i} \\
\text{where} &\quad d_{1,i} = \overline{W_{i\cdot}} - \mathrm{E}\,W_i, d_{2,i} = \overline{W_i^2} - \mathrm{E}\,W_i^2,
\end{aligned}$$

resulting in a rational expression in the central moments of W_i.

□

Proof of Corollaries 5 and 6. Under the corollary assumptions that the class probabilities are uniformly separated from 0 and 1, Y and R are uniformly bounded, implying Assumptions 2 and 3 of Proposition 3. Assumptions 4 of Proposition 3 is assumed directly by the corollaries. The formulas for the bias can be obtained by substituting into the general formulas given by Proposition 3 Taylor expansions for R_i and Y_i,

$$R_i = \sqrt{\frac{p_{T,i}p_{C,i}}{p_{T,i}+p_{C,i}-2p_{T,i}p_{C,i}}} + \frac{d_{T,i}p_{C,i}^{3/2}p_{T,i}^{-1/2} + d_{C,i}p_{T,i}^{3/2}p_{C,i}^{-1/2}}{2(p_{T,i}+p_{C,i}-2p_{T,i}p_{C,i})^{3/2}} + d_{T,i}d_{C,i}\frac{3\sqrt{p_{T,i}p_{C,i}}}{4(p_{T,i}+p_{C,i}-2p_{T,i}p_{C,i})^{5/2}}$$

$$+ \frac{d_{T,i}^2\left(\frac{p_{C,i}}{p_{T,i}}\right)^{3/2}(8p_{T,i}p_{C,i}-4p_{T,i}-p_{C,i}) + d_{C,i}^2\left(\frac{p_{T,i}}{p_{C,i}}\right)^{3/2}(8p_{T,i}p_{C,i}-4p_{C,i}-p_{T,i})}{8(p_{T,i}+p_{C,i}-2p_{T,i}p_{C,i})^{5/2}}$$

$$Y_i = \log\left(\frac{p_{T,i}}{p_{C,i}}\right) + \frac{d_{T,i}}{p_{T,i}} - \frac{d_{C,i}}{p_{C,i}} - \frac{d_{T,i}^2}{2p_{T,i}^2} + \frac{d_{C,i}^2}{2p_{C,i}^2}$$

where $\quad d_{T,i} = \overline{p}_{T,i} - p_{T,i}, \quad d_{C,i} = \overline{p}_{C,i} - p_{C,i},$

for the risk ratio, and analogously for the odds ratio. Since $m_i d_{T,i}$ and $m_i d_{T,i}$ are binomial with success probabilities $p_{T,i}$ and $p_{C,i}$ and m_i trials, the above substitution reduces the general formula to a rational expression in the central moments of a binomial. □