

The Radial Plot in Meta Analysis: Approximations and Applications

John Copas*

University of Warwick, UK

and Claudia Lozada Can

École Polytechnique Fédérale de Lausanne, Switzerland

SUMMARY

Fixed effects meta analysis can be thought of as least squares analysis of the radial plot, the plot of standardized treatment effect against precision (reciprocal of the standard deviation) for the studies in a systematic review. For example, the least squares slope through the origin estimates the treatment effect, and a widely used test for publication bias is equivalent to testing the significance of the regression intercept. However, the usual theory assumes that the within-study variances are known, whereas in practice they are estimated. This leads to extra variability in the points of the radial plot which can lead to a marked distortion in inferences derived from these regression calculations. We derive approximations to the sampling properties of the radial plot and suggest bias corrections to some of the commonly used methods of meta analysis. A simulation study suggests that these bias corrections are effective in controlling significance levels of tests and coverage of confidence intervals. A clinical trials example from the Cochrane Database is used as an illustration.

KEY WORDS: Meta analysis; Radial plot; Bias corrections; Publication bias; Egger test.

*Corresponding author: jbc@stats.warwick.ac.uk; phone 02476523370; fax 02476524532

1 Introduction: radial plots

1.1 The conventional fixed effects model

The standard fixed effects model in meta analysis is that we have k separate studies, each reporting an estimate $\hat{\theta}$ of a common parameter θ . Each estimate (typically a treatment effect) is assumed to be independent and normally distributed

$$\hat{\theta} \sim N\left(\theta, \frac{\sigma^2}{n}\right), \quad (1)$$

where n is the sample size and σ^2 an underlying variance parameter. Note that θ is common (the fixed effects assumption) but n and σ^2 will usually vary across the studies. If we define the standardized treatment effect and the study precision (one over the standard error) as

$$y = \frac{\hat{\theta}\sqrt{n}}{\sigma} \quad \text{and} \quad x = \frac{\sqrt{n}}{\sigma}, \quad (2)$$

then (1) can be written as a linear regression, the so-called *Radial Plot* (Galbraith, 1988)

$$y = \alpha + \theta x + \epsilon, \quad (3)$$

where $\alpha = 0$ and ϵ is a standard normal residual. Ideally, the plot of standardized effects against study precision should be a straight line radiating out from the origin, with slope equal to the true value of θ and known residual variance equal to one. See Sutton *et al.* (2000) for a good introduction to meta analysis and radial plots.

Most of the standard methods of meta analysis can be thought of as regression-like calculations on the radial plot. For example, the usual fixed effects estimate of θ (and its confidence interval) are equivalent to fitting a line through the origin by least squares, the "Q-statistic" used for assessing heterogeneity and defining random effects is the residual sum of squares of this line, and the "Egger test" (Egger *et al.*, 1997) used for assessing publication bias is equivalent to fitting an unconstrained least squares line and testing significance of its intercept. These, and others, are set out in the following list. To simplify the notation here and throughout the paper, we define for any paired sequence (a_i, b_i) , $i = 1, 2, \dots, k$,

$$\bar{a} = \frac{1}{k} \sum a_i, \quad \bar{b} = \frac{1}{k} \sum b_i, \quad s_{ab} = \frac{1}{k} \sum a_i b_i, \quad c_{ab} = \frac{1}{k} \sum (a_i - \bar{a})(b_i - \bar{b}).$$

We are interested in

1. the fixed effects (maximum likelihood) estimate

$$\tilde{\theta} = \frac{s_{xy}}{s_{xx}} \sim N\left(\theta, (k s_{xx})^{-1}\right) \quad (4)$$

2. the confidence interval for θ , given by inverting the pivot

$$Z_1 = (k s_{xx})^{\frac{1}{2}}(\tilde{\theta} - \theta) \sim N(0, 1) \quad (5)$$

3. the Q-statistic (residual sum of squares)

$$Q = s_{yy} - \frac{ks_{xy}^2}{s_{xx}} \sim \chi_{k-1}^2 \quad (6)$$

4. the test statistic for the intercept (Egger test)

$$Z_2 = \left(\frac{kc_{xx}}{s_{xx}} \right)^{\frac{1}{2}} \check{\alpha} \sim N(0, 1) \quad (7)$$

5. the unconstrained test of treatment effect

$$Z_3 = (kc_{xx})^{\frac{1}{2}} \check{\theta} \sim_{(\theta=0)} N(0, 1). \quad (8)$$

Items 4 and 5 are based on the unconstrained least squares estimates

$$\check{\theta} = \frac{c_{xy}}{c_{xx}}, \quad \check{\alpha} = \bar{y} - \check{\theta}\bar{x}.$$

Items 1 to 4 are routine procedures, as mentioned. Item 5 is based on Copas and Malley (2007) who argue that for testing the treatment effect this is more robust than the standard test using Z_1 , robust in the sense of being approximately invariant to selection functions which allow the probability that a study is published to depend on the significance being claimed by that study. The technical argument is that, if $\theta = 0$, the observed y_i s are $N(0, 1)$, and so remain independent and identically distributed even if publication depends in some arbitrary way on these values. The mean of the observed y_i s is no longer zero but this is allowed for, at least approximately, by removing the zero constraint on the intercept.

1.2 The actual model

In practice, the study-specific variances σ^2 will never be known, and so we have to use estimates $\hat{\sigma}^2$ when plotting radial plots and calculating these various quantities. From now on we replace (2) by

$$y = \frac{\hat{\theta}\sqrt{n}}{\hat{\sigma}} \quad \text{and} \quad x = \frac{\sqrt{n}}{\hat{\sigma}}, \quad (9)$$

Replacing σ by $\hat{\sigma}$ in x and y upsets the assumptions necessary for the linear regression (3). The stochastic error in x may lead to bias in the parameter estimates, and the variability in $\hat{\sigma}$ may also induce a within-study correlation between x and y . Both effects mean that the distributions stated in (4) to (8) are no longer valid. Although this problem has been widely acknowledged, little work seems to have been done to explore its consequences. An exception is the Egger test, where several recent simulation studies (Macaskill *et al.*, 2001; Schwarzer *et al.*, 2002; Peters *et al.*, 2006; Harbord *et al.*, 2006) have shown that the true significance level can be substantially inflated, meaning that the test rejects the null hypothesis more often than it should. The effect can be quite marked, even when the individual study sample sizes are reasonably large. Similar biases may also affect other routine methods of meta analysis.

In this paper we study the distribution of x and y in (9), and derive asymptotic approximations to sampling properties of the radial plot, asymptotic in the sense that the study-specific sample sizes are large. This leads to bias corrections to the various statistics listed above, and hence to improved control of significance levels of tests and coverage of confidence intervals. As we shall see, the extent of these biases depends principally on the spread in the values of x across the studies, and on the sampling properties of the particular within-study estimates $\hat{\theta}$ and $\hat{\sigma}^2$. We discuss two contrasting examples:

Example 1. Here $\hat{\theta}$ is the sample mean of a random sample of size n from $N(\theta, \sigma^2)$ and $\hat{\sigma}^2$ is their sample variance. In this case, $\hat{\theta}$ and $\hat{\sigma}^2$ are independent which simplifies much of the theory developed below.

Example 2. This is (by far) the commonest set-up in meta analysis, when we have two treatments and a binary outcome, and so each study has two independent binomial distributions $f_1 \sim \text{bin}(m_1, p_1)$ and $f_2 \sim \text{bin}(m_2, p_2)$. Each study gives an estimated log-odds ratio $\hat{\theta}$ with asymptotic variance σ^2/n based on a total sample size $n = m_1 + m_2$, where

$$\theta = \log \frac{p_1(1-p_2)}{(1-p_1)p_2}, \quad \sigma^2 = \gamma_1 + \gamma_2, \quad (10)$$

and

$$\gamma_j = \frac{1}{\alpha_j p_j (1-p_j)}, \quad \alpha_j = \frac{m_j}{n}; \quad j = 1, 2. \quad (11)$$

In this case, the usual estimates of θ and σ^2 can be substantially correlated.

As an illustration of *Example 2*, Table 1 reports the results of 11 clinical trials into the effectiveness of iron supplementation in pregnancy. This example was used by Schwarzer *et al.* (2007), see Pena-Rosas and Viteri (2006) for full details of this meta analysis from the Cochrane database. The entries in Table 1 are the observed frequencies. For example in the first trial the number of adverse events (low haemoglobin level in late pregnancy) was 0 out of 30 patients on treatment and 14 out of 25 patients on control. The data suggest that iron supplementation has a very strong beneficial effect, the conventional estimate of the log-odds ratio is $\tilde{\theta} = -1.906$ with standard error 0.191. These figures are calculated using the standard convention of adding 0.5 onto each of the observed frequencies to avoid problems with sparse data. Figure 1 is the radial plot for these data. The fit of the fixed effects model is confirmed by the three parallel straight lines drawn on this graph, $y = -1.906x$ (solid line) and $y = -1.906x \pm 1.96$ (dotted lines). Ten out of the eleven points lie between the outer lines, consistent with the point-wise 95% coverage probabilities implied by (3).

Trial	f_1	m_1	f_2	m_2	n
1	0	30	14	25	55
2	0	100	10	107	207
3	2	81	25	84	165
4	17	90	54	95	185
5	7	99	20	98	197
6	1	22	7	23	45
7	3	60	20	60	120
8	4	55	17	54	109
9	3	80	6	44	124
10	1	94	23	55	149
11	0	48	16	42	90

Table 1. Iron supplementation meta analysis.

The bias in the conventional methods is illustrated in Figure 2, which shows the sampling distribution of the pivot Z_1 obtained from a bootstrap simulation. First we fit the fixed effects model to the data in Table 1 by estimating θ as $\tilde{\theta} = -1.906$ and then estimating (p_1, p_2) in each trial by maximum likelihood constrained to have -1.906 as their common log-odds ratio. We have then simulated 2×2 tables with the same sample sizes as in the data. Each set of tables gives a new estimate $\tilde{\theta}$ and hence a value of Z_1 taking the true parameter as $\theta = -1.906$. The dashed line is the kernel density estimate of Z_1 based on 10,000 simulations, which is clearly different from the nominal standard normal distribution shown as the solid line in Figure 2. Notice that the distortion is essentially a shift, there is very little change in variance, suggesting that a bias correction is all that is needed to adjust the pivot before it is inverted to give the confidence interval for θ . The vertical line in Figure 2 indicates the value of $E(Z_1)$ from the asymptotic formula to be developed later in Section 4 below. We see that subtracting this bias from Z_1 successfully restores the distribution to $N(0, 1)$, at least approximately. We will discuss this example further in Section 4.

The main aim of our paper is to show how bias corrections such as this can be made to standard methods of meta analysis. In Section 2 we discuss asymptotic properties of x and y for a single study, and then use these results to study asymptotic properties of the radial plot in Section 3. This leads to the bias corrections of Section 4. In each of these sections we illustrate the general set-up by looking at *Examples 1* and *2*. Section 5 reports a limited simulation study into the sampling properties of some of these bias corrected statistics. Concluding comments are given in Section 6.

2 Asymptotic approximations for a single study

We now consider the joint distribution of $\hat{\theta}$ and $\hat{\sigma}^2$ for a single study. Model (1) states that $\hat{\theta}$ is asymptotically $N(\theta, \sigma^2/n)$; first we extend this to suppose that $(\hat{\theta}, \hat{\sigma}^2)$ is asymptotically jointly normal with mean (θ, σ^2) . We assume that the biases in both estimates are of order $O(n^{-1})$, that $\text{Var}(\hat{\theta}) = \sigma^2/n + O(n^{-2})$, and that $\text{Var}(\hat{\sigma}^2) = O(n^{-1})$. These are standard

properties of maximum likelihood estimates; in practice they will hold for any ‘sensible’ estimates of θ and σ^2 .

We will be interested in the following three functions of x and y as we have re-defined them in (9):

$$u = x - \frac{\sqrt{n}}{\sigma} = \sqrt{n} \left(\frac{1}{\hat{\sigma}} - \frac{1}{\sigma} \right), \quad v = y - \frac{\theta\sqrt{n}}{\sigma} = \sqrt{n} \left(\frac{\hat{\theta}}{\hat{\sigma}} - \frac{\theta}{\sigma} \right), \quad w = v - \theta u = \frac{(\hat{\theta} - \theta)\sqrt{n}}{\hat{\sigma}}. \quad (12)$$

Our approximations will involve two moments in particular, $E(w)$ and $E(uw)$, which from the above assumptions have orders of magnitude

$$E(w) = O(n^{-\frac{1}{2}}) \quad \text{and} \quad E(uw) = O(1). \quad (13)$$

We will also need to note that, again as a consequence of the above assumptions,

$$E(u) = O(n^{-\frac{1}{2}}), \quad \text{Var}(w) = 1 + O(n^{-1}) \quad \text{and} \quad \text{Cov}(w, uw) = O(n^{-\frac{1}{2}}). \quad (14)$$

Example 1 (cont.) For the normal mean example, $\hat{\theta}$ and $\hat{\sigma}^2$ are independent, and $E(\hat{\theta} - \theta) = 0$. It follows immediately that $E(w) = E(uw) = 0$.

Example 2 (cont.) For the 2×2 table, first consider the case of a single binomial distribution, with frequency $f \sim \text{bin}(m, p)$, say. Let z be the asymptotic (large m) standard normal deviate corresponding to f , that is

$$z = \frac{f - mp}{\sqrt{m/\gamma}} \quad \text{and} \quad \gamma = \frac{1}{p(1-p)}.$$

In estimating the log odds, $\log\{p/(1-p)\}$, we follow the standard practice of adding $\frac{1}{2}$ to each of the observed frequencies f and $m - f$, making the sample size effectively $(m + 1)$. Then after some tedious but straightforward algebra we find

$$\log \frac{f + \frac{1}{2}}{m - f + \frac{1}{2}} = \log \frac{p}{1-p} + \gamma^{\frac{1}{2}} z m^{-\frac{1}{2}} + \frac{1}{2} \gamma (1 - 2p)(1 - z^2) m^{-1} + O_p(m^{-3/2}), \quad (15)$$

$$\frac{m + 1}{(f + \frac{1}{2})(m - f + \frac{1}{2})} = \gamma m^{-1} - \gamma^{3/2} (1 - 2p) z m^{-3/2} + O_p(m^{-2}). \quad (16)$$

Extending this to two binomial distributions $f_1 \sim \text{bin}(m_1, p_1)$ and $f_2 \sim \text{bin}(m_2, p_2)$, with $n = m_1 + m_2$, our estimates of θ and σ^2 in (10) are

$$\hat{\theta} = \log \frac{f_1 + \frac{1}{2}}{m_1 - f_1 + \frac{1}{2}} - \log \frac{f_2 + \frac{1}{2}}{m_2 - f_2 + \frac{1}{2}}, \quad (17)$$

$$\hat{\sigma}^2 = n \left\{ \frac{m_1 + 1}{(f_1 + \frac{1}{2})(m_1 - f_1 + \frac{1}{2})} + \frac{m_2 + 1}{(f_2 + \frac{1}{2})(m_2 - f_2 + \frac{1}{2})} \right\}. \quad (18)$$

We can now use (15) and (16) to expand $\hat{\theta}$ and $\hat{\sigma}^2$ in powers of $n^{-\frac{1}{2}}$ and in terms of two independent standard normal deviates z_1 and z_2 . Recalling the notation γ_j in (11), this leads to the following approximations to the quantities u and w in (12),

$$u = \frac{1}{2}(\gamma_1 + \gamma_2)^{-3/2} \{ \gamma_1^{3/2}(1 - 2p_1)z_1 + \gamma_2^{3/2}(1 - 2p_2)z_2 \} + O_p(n^{-\frac{1}{2}}),$$

and

$$\begin{aligned} w &= (\gamma_1 + \gamma_2)^{-\frac{1}{2}} (\gamma_1^{\frac{1}{2}} z_1 + \gamma_2^{\frac{1}{2}} z_2) + \frac{1}{2} n^{-\frac{1}{2}} \left[2u(\gamma_1^{\frac{1}{2}} z_1 - \gamma_2^{\frac{1}{2}} z_2) \right. \\ &+ \left. (\gamma_1 + \gamma_2)^{-\frac{1}{2}} \{ \gamma_1(1 - 2p_1)(1 - z_1^2) - \gamma_2(1 - 2p_2)(1 - z_2^2) \} \right] + O_p(n^{-1}). \end{aligned}$$

These give

$$E(w) = \frac{1}{2}(\gamma_1 + \gamma_2)^{-3/2} \{ \gamma_1^2(1 - 2p_1) - \gamma_2^2(1 - 2p_2) \} n^{-\frac{1}{2}} + O(n^{-1}), \quad (19)$$

$$E(uw) = (\gamma_1 + \gamma_2)^{-\frac{1}{2}} \{ n^{\frac{1}{2}} E(w) \} + O(n^{-\frac{1}{2}}). \quad (20)$$

3 Asymptotic properties of the radial plot

We now consider the expected values of the quantities (4) to (8) when these are calculated from the radial plot co-ordinates (9). Our aim is to derive asymptotic approximations which are valid on the assumption that all the study sample sizes n_1, n_2, \dots, n_k are large. To make this clear, let $N = \sum_1^k n_i$ be the grand total sample size, and $\lambda_i = n_i/N$ be the proportions in each study. Then we imagine that $N \rightarrow \infty$ while the λ_i s remain fixed.

For each study, let

$$a = \frac{\lambda^{\frac{1}{2}}}{\sigma},$$

so that

$$x = \frac{\sqrt{n}}{\hat{\sigma}} = \frac{\lambda^{\frac{1}{2}} N^{\frac{1}{2}}}{\sigma} + u = aN^{\frac{1}{2}} + u,$$

and

$$y = \frac{\hat{\theta}\sqrt{n}}{\hat{\sigma}} = \frac{\theta\lambda^{\frac{1}{2}} N^{\frac{1}{2}}}{\sigma} + v = a\theta N^{\frac{1}{2}} + v.$$

Then

$$s_{xx} = Ns_{aa} + 2N^{\frac{1}{2}}s_{au} + s_{uu} \quad (21)$$

and

$$\begin{aligned} s_{xy} &= \theta Ns_{aa} + N^{\frac{1}{2}}(s_{av} + \theta s_{au}) + s_{uv} \\ &= \theta s_{xx} + N^{\frac{1}{2}}s_{av} + s_{uv}, \end{aligned} \quad (22)$$

where, as before, $w = v - \theta u$. Thus the fixed effects estimate of θ is

$$\tilde{\theta} = \frac{s_{xy}}{s_{xx}} = \theta + \frac{s_{av}}{s_{aa}} N^{-\frac{1}{2}} + O_p(N^{-1}). \quad (23)$$

Similarly, the pivot in (5) comes to

$$Z_1 = \left(\frac{k}{s_{aa}} \right)^{\frac{1}{2}} \left\{ s_{av} + \left(s_{uv} - \frac{s_{av}s_{au}}{s_{aa}} \right) N^{-\frac{1}{2}} \right\} + O_p(N^{-1}), \quad (24)$$

and the residual sum of squares in (6) is

$$Q = k \left\{ s_{ww} - \frac{s_{wa}^2}{s_{aa}} - \frac{2s_{wa}}{s_{aa}^2} (s_{aa}s_{uw} - s_{au}s_{aw}) N^{-\frac{1}{2}} \right\} + O_p(N^{-1}). \quad (25)$$

Now, for each study, let

$$b = \lim_{N \rightarrow \infty} N^{\frac{1}{2}} E(w) \quad \text{and} \quad c = \lim_{N \rightarrow \infty} E(uw).$$

From(13), all three study-specific quantities (a, b, c) are of order $O(1)$ in our approximations as $N \rightarrow \infty$. Then the expectations of the random quantities appearing in (23) to (25) are

$$E(s_{aw}) = s_{ab}N^{-\frac{1}{2}} + O(N^{-1}), \quad E(s_{uw}) = \bar{c} + O(N^{-\frac{1}{2}}), \quad E(s_{au}s_{aw}) = k^{-1}s_{a^2c} + O(N^{-\frac{1}{2}}),$$

and

$$E(s_{ww}) = 1 + O(N^{-1}), \quad E(s_{wa}^2) = k^{-1}s_{aa} + O(N^{-1}).$$

Thus

$$\begin{aligned} E(\tilde{\theta}) &= \theta + O(N^{-1}), \\ E(Z_1) &= \left(\frac{k}{s_{aa}} \right)^{\frac{1}{2}} \left\{ s_{ab} + \bar{c} - \frac{s_{a^2c}}{ks_{aa}} \right\} N^{-\frac{1}{2}} + O(N^{-1}) \end{aligned} \quad (26)$$

and

$$E(Q) = k - 1 + O(N^{-1}).$$

For Z_2 and Z_3 in (7) and (8) we need the corrected sums of squares and products c_{xx} and c_{xy} . Expressions for these are exactly the same as (21) and (22) but with the obvious replacement of terms like s_{ab} by the corresponding covariance terms c_{ab} . Reworking the calculation of (26) gives

$$E(Z_3) = \left(\frac{k}{c_{aa}} \right)^{\frac{1}{2}} \left\{ c_{ab} + \left(1 - \frac{2}{k} \right) \bar{c} - \frac{c_{(a-\bar{a})^2c}}{kc_{aa}} \right\} N^{-\frac{1}{2}} + O(N^{-1}). \quad (27)$$

For Z_2 we also need $\bar{x} = \bar{a}N^{\frac{1}{2}} + \bar{u}$ and $\bar{y} = \theta\bar{a}N^{\frac{1}{2}} + \bar{v}$, and so the least squares intercept becomes

$$\begin{aligned} \check{\alpha} &= \bar{y} - \check{\theta}\bar{x} \\ &= \bar{w} - \frac{c_{a,w}\bar{a}}{c_{a,a}} - \left\{ \frac{c_{a,w}\bar{u}}{c_{a,a}} + \frac{\bar{a}(c_{u,w}c_{a,a} - 2c_{a,w}c_{a,u})}{c_{a,a}^2} \right\} N^{-\frac{1}{2}} + O_p(N^{-1}). \end{aligned}$$

Scaling by the normalizing factor as in (7) and taking expectations leads to

$$E(Z_2) = \left(\frac{k}{c_{aa}s_{aa}^3} \right)^{\frac{1}{2}} AN^{-\frac{1}{2}}, \quad (28)$$

where

$$A = s_{aa}(c_{aa}\bar{b} - c_{ab}\bar{a}) - \frac{1}{k} \left\{ c_{ac}(s_{aa} - 2\bar{a}^2) + (k-2)s_{aa}\bar{a}\bar{c} + \frac{\bar{a}c_{(a-\bar{a})^2c}}{c_{aa}}(2s_{aa} - \bar{a}^2) \right\}. \quad (29)$$

Example 1 (cont.) Here $b = c = 0$ and so the leading bias terms of order $O(N^{-\frac{1}{2}})$ in all of the above expressions are zero. Although these estimates and pivots are not exactly unbiased, their biases are an order of magnitude smaller than in the general case.

Example 2 (cont.) The bias terms involve means and mean squares of b and c . From (19) and (20), c and b are

$$c = \lim_{N \rightarrow \infty} E(uw) = \frac{\gamma_1^2(1 - 2p_1) - \gamma_2^2(1 - 2p_2)}{2(\gamma_1 + \gamma_2)^2} \quad (30)$$

and

$$b = \lim_{N \rightarrow \infty} N^{\frac{1}{2}} E(w) = \frac{n^{\frac{1}{2}} E(w)}{\lambda^{\frac{1}{2}}} = \frac{n^{\frac{1}{2}} E(w)}{a\sigma} = \frac{c}{a}, \quad (31)$$

this last step following from (20) as $\sigma^2 = \gamma_1 + \gamma_2$.

4 Bias corrections

From (24) we have

$$\text{Var}(Z_1) = \left(\frac{k}{s_{aa}} \right) \left(\text{Var}(s_{aw}) + 2N^{-\frac{1}{2}} \text{Cov}(s_{aw}, B) \right) + O(N^{-1}),$$

where B is the factor multiplying $N^{-\frac{1}{2}}$ in expression (24). Now B can be written as a linear combination of products of the form $u_i w_j$ and so from (14) we have

$$\text{Var}(s_{aw}) = \frac{s_{aa}}{k} + O(N^{-1}) \quad \text{and} \quad \text{Cov}(s_{aw}, B) = O(N^{-\frac{1}{2}}).$$

It follows that

$$\text{Var}(Z_1) = 1 + O(N^{-1}),$$

to be compared with (26) which establishes that $E(Z_1) = O(N^{-\frac{1}{2}})$. This confirms the impression from Figure 2 that the main difference between $N(0, 1)$ and the actual distribution of Z_1 is in the mean rather than the variance.

If we ignore terms of order $O(N^{-1})$, a normal approximation for Z_1 gives the adjusted confidence interval for θ with limits

$$\tilde{\theta} - (ks_{xx})^{-\frac{1}{2}} \{E(Z_1) \pm z_\alpha\}, \quad (32)$$

where $E(Z_1)$ is the bias of order $O(N^{-\frac{1}{2}})$ given by (26), and z_α is the standard normal percentage point needed to achieve the desired coverage $(1 - \alpha)$. The conventional confidence interval is just (32) with the bias term omitted. The analogous two-sided P-value for treatment effect is

$$P = 2\Phi\{-|(ks_{xx})^{\frac{1}{2}}\tilde{\theta} - E(Z_1)|\},$$

where Φ is the standard normal cumulative distribution function. Similarly, the adjusted two-sided P-values for the tests based on Z_2 and Z_3 are

$$2\Phi\{-|Z_2 - E(Z_2)|\} \quad \text{and} \quad 2\Phi\{-|Z_3 - E(Z_3)|\}. \quad (33)$$

To implement these adjustments in practice we have to estimate the bias terms using the formulae given in the last section, the details depending on the statistical characteristics of the individual studies.

Example 1 (cont.) No bias corrections are needed in this case.

Example 2 (cont.) From (10),

$$aN^{\frac{1}{2}} = \frac{\sqrt{n}}{\sigma} = \sqrt{\frac{n}{\gamma_1 + \gamma_2}}.$$

This, together with (30) and (31), shows that (a, b, c) can all be written as functions of the values of (γ_1, γ_2) and hence of (p_1, p_2) in the k studies. To estimate (p_1, p_2) we should exploit the fixed effects assumption, that the pairs (p_1, p_2) are related through the common log-odds ratio θ . The constrained maximum likelihood estimates of p_1 and p_2 given that

$$\Theta(p_1, p_2) = \log \frac{p_1(1-p_2)}{(1-p_1)p_2} = \theta$$

are

$$p_1(\lambda) = \frac{f_1 + \lambda + \frac{1}{2}}{m_1 + 1} \quad \text{and} \quad p_2(\lambda) = \frac{f_2 - \lambda + \frac{1}{2}}{m_2 + 1}, \quad (34)$$

where the Lagrange multiplier λ is a solution of the quadratic equation $\Theta\{p_1(\lambda), p_2(\lambda)\} = \theta$. For consistency with the earlier definitions of x and y , we have again added one half onto all of the observed frequencies in these calculations. By examining the function $\Theta(p_1, p_2)$ it is easy to check that the quadratic equation for λ has two real solutions, and that it is uniquely the larger solution which gives values of p_1 and p_2 in $[0, 1]$. We suggest taking $\theta = \tilde{\theta} = \sum xy / \sum x^2$, the standard fixed effects estimate of θ , finding the corresponding pairs (p_1, p_2) , and then using these to calculate (a, b, c) .

To set out the calculation more explicitly, we have k studies each with data $f_1 \sim \text{bin}(m_1, p_1)$ and $f_2 \sim \text{bin}(m_2, p_2)$. The steps are

- First calculate the study-specific estimates $(\hat{\theta}, \hat{\sigma}^2)$ in (17) and (18), and hence the usual radial plot co-ordinates (x, y) in (9) and the appropriate radial plot regression statistics.
- Let $\eta = \exp(\sum xy / \sum x^2)$. If $\eta \neq 1$, calculate the Lagrange multipliers

$$\begin{aligned} \lambda &= \{2(1-\eta)\}^{-1} [-\{f_1 + m_2 - f_2 + 1 + \eta(f_2 + m_1 - f_1 + 1)\} \\ &\quad + \{\{f_1 + f_2 - m_2 + \eta(m_1 - f_1 - f_2)\}^2 + 4\eta(m_1 + 1)(m_2 + 1)\}^{\frac{1}{2}}], \end{aligned}$$

and hence each study's estimate of (p_1, p_2) in (34). If $\eta = 1$ (no treatment effect) then we simply take $p_1 = p_2 = (f_1 + f_2 + 1)/(m_1 + m_2 + 2)$. We can now calculate the estimates $(\hat{\gamma}_1, \hat{\gamma}_2)$ from (11) and hence the following four quantities for each study:

$$\begin{aligned} a &= N^{-\frac{1}{2}} \sqrt{\frac{n}{\hat{\gamma}_1 + \hat{\gamma}_2}}, \\ c &= \frac{\hat{\gamma}_1^2(1 - 2\hat{p}_1) - \hat{\gamma}_2^2(1 - 2\hat{p}_2)}{2(\hat{\gamma}_1 + \hat{\gamma}_2)^2}, \\ b &= \frac{c}{a}, \\ d &= (a - \bar{a})^2. \end{aligned}$$

- Calculate the average values across the k studies of the quantities (a, b, c, d) and their appropriate empirical mean squares and covariances, and substitute into the relevant bias formula needed for the confidence interval or test of interest.

To illustrate these calculations we return to the medical example mentioned in Section 1, with data set out in Table 1. First we calculate the radial plot statistics from the points in Figure 1. We then fit the fixed effects model to the observed 2×2 tables by estimating (p_1, p_2) as explained above, and hence estimates of (a, b, c, d) for each study, and hence the bias corrections for Z_1 and Z_2 . The null hypothesis being tested by Z_3 is $H_0 : \theta = 0$, and so to estimate the bias correction for Z_3 we need to fit the null model instead ($\eta = 1$ in the above notation). The results are as follows:

Confidence interval for θ . The conventional estimate of the log-odds ratio is $\tilde{\theta} = -1.906$ with standard error 0.191. This gives the 95% confidence interval

$$-1.906 \pm 1.96 \times 0.191 = (-2.280, -1.532).$$

The estimate of $E(Z_1)$ in (26) is +0.879 giving the corrected confidence interval (32)

$$-1.906 - 0.191 \times (0.879 \pm 1.96) = (-2.447, -1.700).$$

In this case $\tilde{\theta}$ is sufficiently accurate that the bias correction results in only a small shift in the values of θ , but the coverage properties of the conventional interval would be very poor (the bootstrap distribution in Figure 2 suggests that the actual coverage would be nearer 85% than 95%).

Intercept test for publication bias. The value of (7) is $Z_2 = -2.844$ giving the conventional P-value as $P = 0.004$, highly significant evidence for publication bias. The estimate of $E(Z_2)$ from (28) and (29) comes to -0.927 , and so the bias adjusted P-value in (33) is $P = 0.055$, suggesting that the real evidence is substantially weaker than the naive analysis suggests. This agrees with the analysis of this example in Schwarzer *et al.* (2007), who develop a completely different test for publication bias based on the exact conditional distributions of the 2×2 tables, and again conclude that the evidence for publication bias provided by the conventional intercept test is much exaggerated.

Unconstrained test of treatment effect. The evidence for a non-zero treatment effect is extremely strong when judged by the confidence intervals discussed above, with or without the bias correction on Z_1 . However the intercept test raises doubts about publication bias, even with the bias correction on Z_2 the P-value only just exceeds 5%. It is well known that if there is a selection effect in meta analysis then to ignore it can be extremely misleading. Thus it might be thought safer in this case to allow for the possibility of a selection effect, and use the robust test based on Z_3 . The value of (8) comes to $Z_3 = -1.698$, giving a nominal P-value of $P = 0.090$: the evidence for the treatment effect is now much weaker. The bias correction in (27) under the null model is -0.0818 , adjusting the P-value to $P = 0.106$. Here the bias correction is unimportant, but allowing for publication bias in this way has been catastrophic as far as the conclusion of this particular meta analysis is concerned.

5 Simulation study

In this section we give a very brief summary of some simulation results for 2×2 tables. Clearly it is impossible to cover all possibilities: we illustrate our results by a few representative cases and suggest some qualitative conclusions.

For Figures 3 and 4, and Table 2, we have taken $k = 50$, generated random sample sizes m_1 and m_2 uniformly between 150 and 300, and defined p_E , the average of p_1 and p_2 on the logit scale, to be 0.3. For any true value of θ we can then generate random 2×2 tables, and for each set of tables calculate the statistics of interest and estimate their bias corrections as set out in Section 4. The simulations are repeated 100,000 times.

Figure 3 looks at Z_1 with $\theta = \log(0.2)$ and shows kernel density estimates of Z_1 and the corresponding estimates of $Z_1 - E(Z_1)$. The distribution of $Z_1 - E(Z_1)$ is virtually indistinguishable from the standard normal, but the distribution of Z_1 is noticeably shifted to the right, in the same direction as the bias noted earlier for the example in Figure 2. In Figure 3, the estimates of $E(Z_1)$ have been very effective at removing the bias.

Figure 4 takes $\theta = \log(0.5)$ and shows the kernel density estimates of Z_2 and the estimated $Z_2 - E(Z_2)$. Here there is an even larger shift, this time to the left, but again the distribution of $Z_2 - E(Z_2)$ is virtually indistinguishable from the standard normal.

The short vertical lines on these figures indicate the 5th and 95th percentiles of the standard normal distribution. Thus the areas to each side of these lines indicate the actual significance levels of tests based on Z_1 and Z_2 , or equivalently the actual coverage of confidence intervals based on Z_1 . The results are summarized in Table 2 for $\theta = \log(0.67)$, $\log(0.5)$ and $\log(0.2)$. For the cases shown, the distortion in significance levels is particularly severe for the intercept test Z_2 , these figures being in line with the findings of Macaskill *et al.* (2001), Schwarzer *et al.* (2002), Peters *et al.* (2006) and Harbord *et al.* (2006).

θ	statistic	left tail %	right tail %	significance level %
log(0.67)	Z_1	3.73	6.12	9.85
	$Z_1 - E(Z_1)$	4.95	4.95	9.90
	Z_2	20.65	0.72	21.37
	$Z_2 - E(Z_2)$	5.08	5.13	10.21
log(0.5)	Z_1	3.15	7.20	10.35
	$Z_1 - E(Z_1)$	5.0	5.0	10.0
	Z_2	38.11	0.12	38.23
	$Z_2 - E(Z_2)$	4.51	5.14	9.65
log(0.2)	Z_1	1.66	11.73	13.39
	$Z_1 - E(Z_1)$	5.05	5.16	10.21
	Z_2	88.07	0	88.07
	$Z_2 - E(Z_2)$	1.84	6.95	8.79

Table 2. Error rates of Z_1 and Z_2 for nominal 10% (two-sided) and 5% (one-sided) tests.

By comparing such simulations with similar results for other configurations we can draw some tentative conclusions about how these biases depend on the particular characteristics of the meta analysis. We find:

- bias tends to increase with k , the number of studies
- bias tends to increase with $|\theta|$ (the stronger the treatment effect)
- bias tends to increase as the average response probability p_E becomes more extreme (nearer 0 or 1)
- bias tends to decrease as the sample sizes m_1 and m_2 increase
- bias tends to increase as the trials become more unbalanced (m_1 different from m_2)
- bias is relatively insensitive to variations in p_1 and p_2 between trials (trials have the same value of θ but different values of p_E)
- on the whole, bias is more marked for the intercept test Z_2 than for the tests or confidence intervals based on Z_1 and Z_3
- the accuracy of the bias estimates deteriorates as the data become more sparse (m_1 and p_1 , or m_2 and p_2 , both small). These are cases, however, where bias can be particularly severe, but the bias corrections can still be worthwhile in the sense of being less misleading than the conventional statistics (this is clearly shown in the last two lines of Table 2 where the Z_2 test is grossly misleading: at this setting there are likely to be several zeros in the data).

6 Comments

1. Each point (9) of the radial plot can be thought of as an estimate of the true point $(\theta\sqrt{n}/\sigma, \sqrt{n}/\sigma)$ for that study. These true points become unbounded as $n \rightarrow \infty$. Thus

the asymptotics in this paper are somewhat non-standard, unlike more usual asymptotic discussions where we are studying properties of estimates of fixed parameters. Note in particular that the covariance notation c_{ab} which we have used throughout the paper refers to variation between rather than within studies. For example, the variance c_{aa} is a positive quantity even though the study specific a 's have been defined as fixed parameters.

2. We have referred to Z_2 as the intercept test rather than the Egger test, to make the technical distinction that Z_2 uses the fact that the residual variance of the radial plot under the fixed effects model is known to be one, whereas the Egger test uses the estimated residual mean square as in standard regression analysis. Although less efficient than Z_2 under fixed effects, the Egger test has the advantage of being more robust to heterogeneity between the studies.

3. As mentioned in the Introduction, the inadequacies of the Egger test have been widely acknowledged. Several recent papers have suggested that the Egger test be abandoned in favour of better tests for publication bias, for example Harbord *et al.* (2005), Macaskill *et al.* (2001), Peters *et al.* (2006) and Shchwarzzer *et al.* (2007). All these tests are claimed to have a more stable Type I error rate than the Egger test. Similarly, exact methods for 2×2 tables give, at least in principle, better confidence intervals and tests for θ . However the conventional methods based on regression-type calculations from the radial plot are very widely used, and our approach is to retain the simplicity of these methods whilst addressing their statistical problems with the proposed bias corrections.

4. We have assumed the fixed effects model throughout. The usual practice is to test this assumption by calculating Q in (6) and if this is significant as χ^2 on $(k - 1)$ degrees of freedom to estimate a between-studies variance τ^2 and use the random effects model instead (Sutton *et al.*, 2000). Essentially, this is equivalent to redefining the radial plot coordinates by replacing σ^2/n by $\tau^2 + \sigma^2/n$. The conventional calculations of $\tilde{\theta}$ and Z_1 take exactly the same form, but the bias $E(Z_1)$ will change. Simulations suggest that when τ^2 is small the size of this bias is quite similar, but we have no theory to back this up. To fully rework the theory of Section 3 for the random effects model would be much more challenging, as estimates of τ^2 depend on all the data and not just on the individual points in the radial plot.

5. The bias formulae for *Example 2* simplify for two special (but important) cases, when the two arms of each trial are balanced ($m_1 = m_2$ in all trials), and for the null hypothesis H_0 ($p_1 = p_2$ in all trials). If both hold (balanced trials with no treatment effect), the biases are all zero since $c = 0$ in (30). In particular, Z_3 is an approximately unbiased test for balanced clinical trials.

REFERENCES

- Copas, J. B. and Malley, P. F. (2007). A robust P-value for treatment effect in meta analysis with publication bias. *Research Report*, Dept of Statistics, University of Warwick.
- Egger, M., Smith, G. D., Schneider, M. and Minder, C. (1997). Bias in meta analysis detected by a simple graphical test. *Brit. Med. J.*, **315**, 629-634.
- Galbraith, R. F. (1988). A note on graphical representation of estimated odds ratios from several clinical trials. *Statist. in Med.*, **7**, 889-894.

- Harbord, R. M., Egger, M. and Sterne, J. A. C. (2006). A modified test for small study effects in meta analysis of controlled trials with binary endpoints. *Statist. in Med.*, **25**, 3443-3457.
- Macaskill, P., Walter, S. D. and Irwig, L. (2001). A comparison of methods to detect publication bias in meta analysis. *Statist. in Med.*, **20**, 641-654.
- Pena-Rosas, J. P. and Viteri, F. E. (2006). Effects of routine oral iron supplementation with or without folic acid for women in pregnancy. *Cochrane Database of Systematic Reviews*, **2006**, **Issue 3**, Art No.: CD004736. DOI: 10.1002/14651858.CD004736.pub2.
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R. and Rushton, L. (2006). Comparison of two methods to detect publication bias in meta analysis. *J. Am. Med. Assoc.*, **295**, 676-680.
- Shwarzer, G., Antes, G. and Shumacher, M. (2002). Inflation in Type I error rate in two statistical tests for the detection of publication bias in meta analysis with binary outcomes. *Statist. in Med.*, **21**, 2465-2477.
- Shwarzer, G., Antes, G. and Shumacher, M. (2007). A test for publication bias in meta-analysis with sparse data. *Statist. in Med.*, **26**, 721-733 .
- Sutton, A. J., Abrams, K. R., Jones, D. R. and Sheldon, T. A. (2000). *Methods for meta-analysis in medical research*. Chichester: Wiley.

Captions for tables and figures

Table 1. Iron supplementation meta analysis.

Table 2. Error rates of Z_1 and Z_2 for nominal 10% (two-sided) and 5% (one-sided) tests.

Figure 1. Radial plot for iron supplementation meta analysis.

Figure 2. Distribution of Z_1 for iron supplementation meta analysis.

Figure 3. Simulation study: distribution of Z_1 .

Figure 4. Simulation study: distribution of Z_2 .

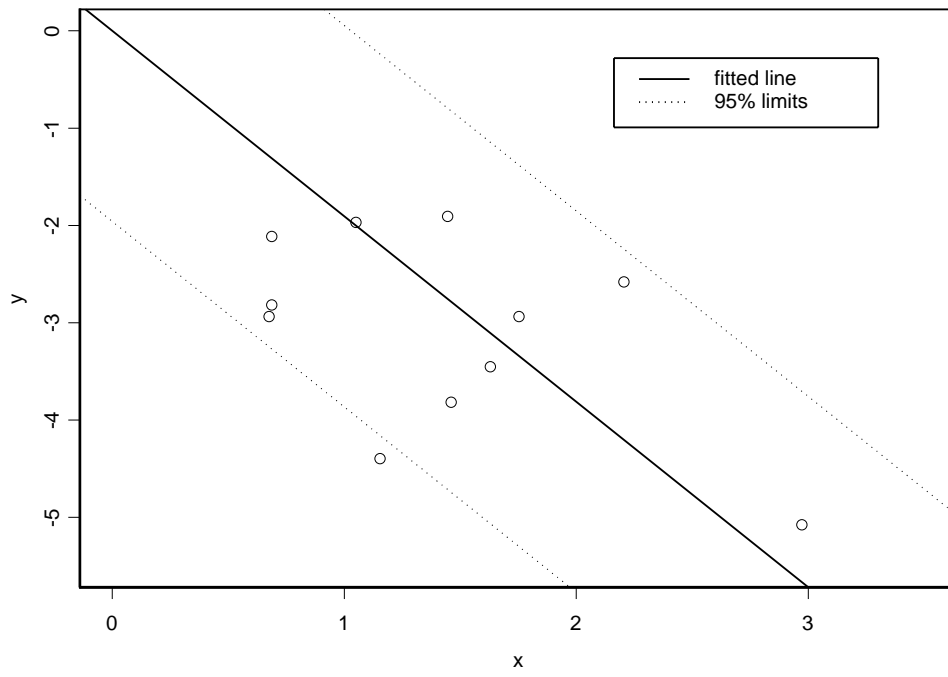


Figure 1: Radial plot for iron supplementation meta analysis.

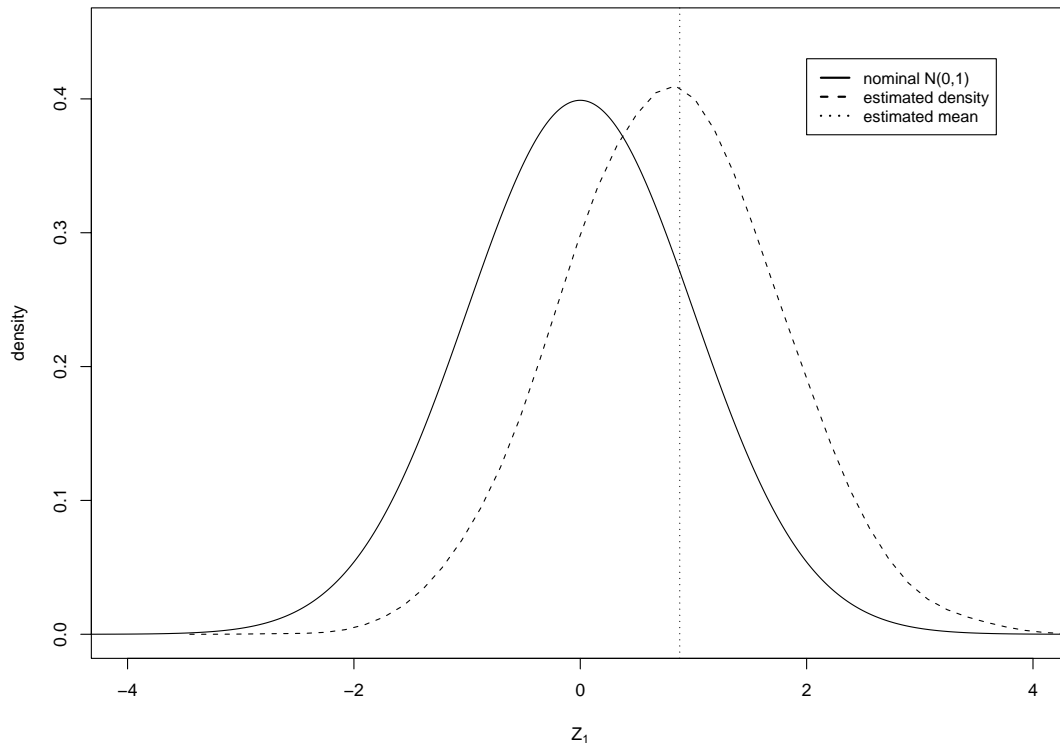


Figure 2: Distribution of Z_1 for iron supplementation meta analysis.

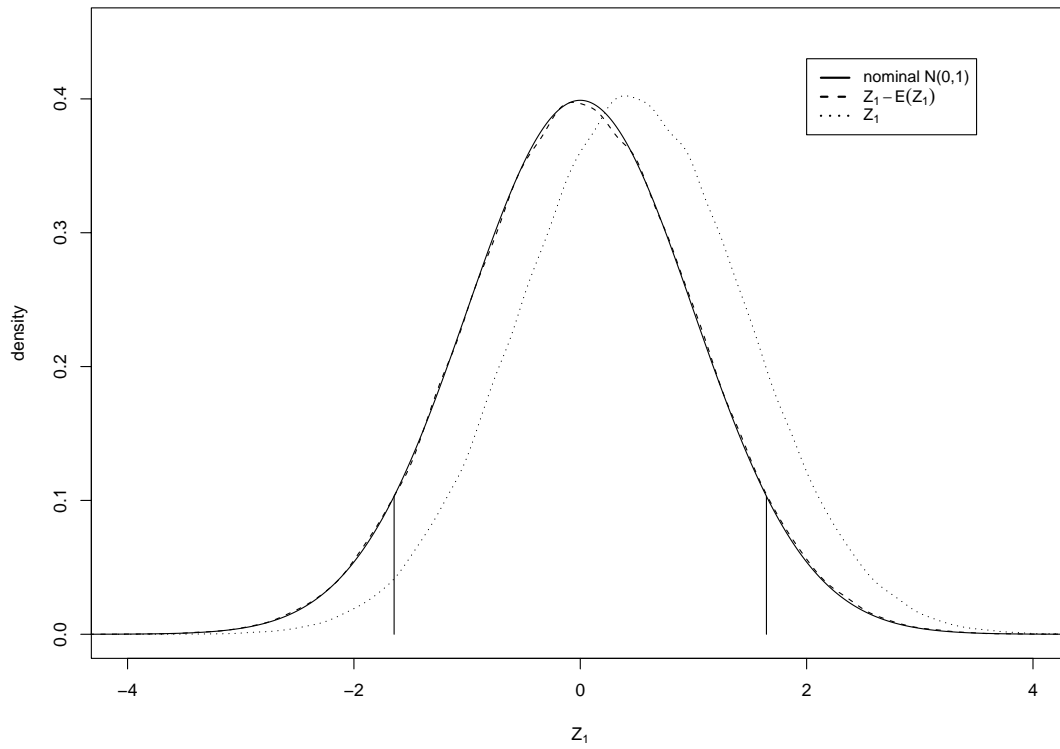


Figure 3: Simulation study: distribution of Z_1 .

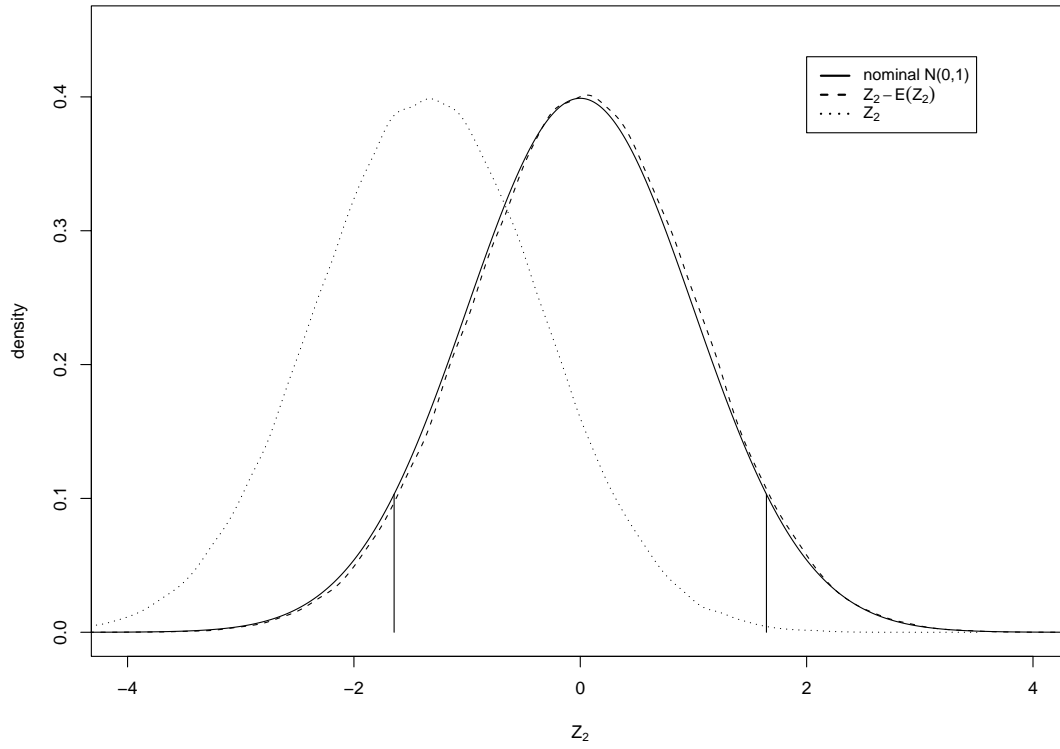


Figure 4: Simulation study: distribution of Z_2 .