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. This is illustrated by a clinical trials example from the Cochrane Database. 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.

Key words: Meta analysis; Radial plot; Bias corrections; Publication bias; Egger test.

*Corresponding author: jbc@stats.warwick.ac.uk; phone 44(0)2476523370; fax 44(0)2476524532

1 Introduction and example

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

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

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. Given $(\hat{\theta}_i, \sigma_i^2, n_i)$ for the k studies, the maximum likelihood estimate of θ, and its variance, are

$$\tilde{\theta} = \frac{\sum n_i \sigma_i^{-2} \hat{\theta}_i}{\sum n_i \sigma_i^{-2}} \quad , \quad \sigma_{\tilde{\theta}}^2 = \frac{1}{\sum n_i \sigma_i^{-2}} \ . \tag{2}$$

Confidence intervals and tests for θ are based on the fact that

$$Z_1 = (\sum n_i \sigma_i^{-2})^{\frac{1}{2}} (\tilde{\theta} - \theta) \tag{3}$$

has a standard normal distribution.

A good way of looking at $(\hat{\theta}_i, \sigma_i^2, n_i)$ in a meta analysis is to plot the data as a *Radial Plot* (Galbraith, 1988). This plots the standardized treatment effect against the precision (one over the standard error) of each study, or y against x where

$$y = \frac{\hat{\theta}\sqrt{n}}{\sigma} \quad \text{and} \quad x = \frac{\sqrt{n}}{\sigma} \ . \tag{4}$$

In terms of the radial plot, (1) is the linear regression model

$$y = \alpha + \theta x + \epsilon, \tag{5}$$

where $\alpha = 0$ and ϵ is a standard normal residual. Thus, if the model is correct, 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. Further, $\tilde{\theta}$ is just the slope of the least squares line through the origin, so (2) and (3) are the familiar regression calculations

$$\tilde{\theta} = \frac{s_{xy}}{s_{xx}} \ , \ \sigma_{\tilde{\theta}}^2 = \frac{1}{k s_{xx}} \ , \ Z_1 = (k s_{xx})^{\frac{1}{2}} (\tilde{\theta} - \theta) \ .$$

In this and throughout the paper we use the generic notation, for any k pairs of numbers (a_i, b_i),

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

See Sutton *et al.* (2000) for a good general introduction to meta analysis and radial plots.

For reasons that will become clear later, we have used a slightly different notation from that usually used in meta analysis. By defining the variance of $\hat{\theta}$ to be σ^2/n rather than simply σ^2, we are making explicit the role of the within-study sample sizes.

As the motivating example for this paper, Table 1 reports the results of 11 clinical trials into the effectiveness of iron supplementation in pregnancy. This example comes from the Cochrane database (Pena-Rosas and Viteri, 2006; Schwarzer *et al.*, 2007). The entries in Table 1 are 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 (f_1 out of m_1) and 14 out of 25 patients on control (f_2 out of m_2). The data clearly suggest that iron supplementation has a very strong beneficial effect. If θ is the underlying log odds ratio measuring the treatment effect, then for each study the usual estimate of θ is

$$\hat{\theta} = \log\left\{\frac{(f_1+0.5)(m_2-f_2+0.5)}{(m_1-f_1+0.5)(f_2+0.5)}\right\}\ , \tag{6}$$

with variance σ^2 taken to be

$$\hat{\sigma}^2 = n\left(\frac{1}{f_1+0.5}+\frac{1}{m_1-f_1+0.5}+\frac{1}{f_2+0.5}+\frac{1}{m_2-f_2+0.5}\right)\ , \tag{7}$$

where n is the total study sample size $n = m_1 + m_2$. In these calculations we have followed the standard convention of adding 0.5 onto each frequency to improve bias and avoid problems with sparse data. Substituting these values into (2) gives the meta analysis estimate of θ to be $\tilde{\theta} = -1.906$ with standard error $\sigma_{\tilde{\theta}} = 0.191$.

To visualize the model, Figure 1 shows 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 (5).

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.

Although the radial plot suggests a reasonable fit of the model, we know that the model cannot be *exactly* correct because we have ignored the fact that (7) is merely an estimate and not the true variance of each study. As mentioned, the validity of the usual analysis depends on the distribution of Z_1 being standard normal. We can check this by estimating the actual distribution of Z_1 through a bootstrap simulation. To do this, we take the

estimate $\tilde{\theta} = -1.906$ as if it was the true value of θ, and estimate the probabilities (p_1, p_2) for the two arms of each trial so that they have -1.906 as their common log-odds ratio (details of how to do this will be discussed later in Section 3). We then simulate sets of 2×2 tables by generating random observations from the binomial distributions $f_1 \sim \text{bin}(m_1, p_1)$ and $f_2 \sim \text{bin}(m_2, p_2)$. Each set of tables gives a new estimate $\tilde{\theta}$ and hence a value of Z_1 with $\theta = -1.906$. The dashed line in Figure 2 shows the kernel density estimate of Z_1 based on $10,000$ simulations. This is clearly different from the nominal standard normal distribution shown as the solid line. The distortion is a substantial shift to the right. The fact that the variance stays about the same suggests that for valid confidence intervals we should adjust the value Z_1 with a bias correction before assuming it is standard normal. The vertical dotted line in Figure 2 indicates the value of $\text{E}(Z_1)$ calculated from the asymptotic formula to be developed later in Sections 2 and 3. We see that subtracting this bias from Z_1 successfully restores the distribution to $N(0, 1)$, at least approximately. The object of our paper is to show how we can improve meta analysis calculations by developing bias corrections of this kind. We will return to this example again in Section 3.

Other standard methods in meta analysis can also be reduced to regression-like calculations on the radial plot. The residual sum of squares from the least squares line through the origin,

$$Q = ks_{yy} - \frac{ks_{xy}^2}{s_{xx}} = \sum_i \frac{n_i}{\sigma_i^2}(\hat{\theta}_i - \tilde{\theta})^2 \ , \tag{8}$$

is just the usual "Q-statistic" for testing heterogeneity between the studies. Under the fixed effects model (1), Q is chi-squared on $(k-1)$ degrees of freedom. A large value of Q indicates that there are systematic differences between the studies — the usual approach is then to use Q to estimate the random effects variance in a random effects model (Sutton *et al.*, 2000, section 5.2). The model assertion that the intercept of the radial plot is zero can also be tested, by fitting an unconstrained least squares line to the radial plot and testing its intercept to give the test statistic

$$Z_2 = \left(\frac{kc_{xx}}{s_{xx}}\right)^{\frac{1}{2}} (\bar{y} - \breve{\theta}\bar{x}) \ , \ \breve{\theta} = \frac{c_{xy}}{c_{xx}} \ . \tag{9}$$

Under the model, Z_2 is standard normal. Testing Z_2 is equivalent to the "Egger test", widely used in meta analysis to test for publication bias or a "small study effect" (Egger *et al.*, 1997). Finally, testing significance of the slope $\breve{\theta}$ of the unconstrained least squares line provides another way of testing the treatment effect (that $\theta \neq 0$). Under the model (1) this is less powerful than the test based on Z_1, but Copas and Malley (2008) show that it is much more robust to publication bias. This gives the robust test statistic

$$Z_3 = (kc_{xx})^{\frac{1}{2}}\breve{\theta} \ , \tag{10}$$

again standard normal under the model.

The example has already highlighted the main cause of bias, that in practice we have to use estimates $\hat{\sigma}^2$ instead of the true study-specific variances σ^2 when plotting radial plots and calculating these various quantities. So from now on we replace (4) by

$$y = \frac{\hat{\theta}\sqrt{n}}{\hat{\sigma}} \quad \text{and} \quad x = \frac{\sqrt{n}}{\hat{\sigma}} \ . \tag{11}$$

Replacing σ by $\hat{\sigma}$ in x and y upsets the assumptions necessary for the linear regression (5). 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 for Z_1, Q, Z_2 and Z_3 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.

In Section 2 of this paper we study the distribution of x and y in (11) and give asymptotic approximations to some of the sampling properties of the radial plot, asymptotic in the sense that the study-specific sample sizes are large. This leads in Section 3 to bias corrections to the various statistics listed above, and hence to improved control of significance levels of tests and coverage of confidence intervals. We also return to the example in Section 3, and then report the results of a simulation study in Section 4. Brief comments and conclusions are listed in Section 5. The Appendix outlines the derivation of some of the formulae used in Sections 2 and 3.

2 Asymptotic theory of radial plots

Since there are k different sample sizes n_i in a meta analysis, we first need to be clear what we mean by "asymptotic". The accuracy of the approximations developed below are stated in terms of the overall sample size $N = \sum_1^k n_i$, assuming that the proportional sample sizes n_i/N are fixed. Our numerical results suggest that in practice these approximations will be useful provided that the sample sizes in the majority of studies are not too small.

Properties of the radial plot coordinates x and y in (11) depend on the statistical properties of $\hat{\theta}$ and $\hat{\sigma}^2$, and hence on the type and design of the studies being combined in the meta analysis. We show in the Appendix that there are three important quantities for each study:

$$a = \left(\frac{n}{N\sigma^2}\right)^{\frac{1}{2}} \ , \ b = \mathrm{E}\left\{\frac{(nN)^{\frac{1}{2}}(\hat{\theta}-\theta)}{\hat{\sigma}}\right\} \ , \ d = \mathrm{E}\left\{\frac{n(\hat{\theta}-\theta)(\hat{\sigma}-\sigma)}{\sigma\hat{\sigma}^2}\right\} \ . \qquad (12)$$

The quantity a is defined directly in terms of n and σ^2. Two special cases of interest for b and d are:

Special case 1: normal data. If, in each study, $\hat{\theta}$ is the sample mean of a random sample of size n from $N(\theta, \sigma^2)$ and $\hat{\sigma}^2$ is the sample variance, then $\hat{\theta}$ and $\hat{\sigma}^2$ are independent and so $b = d = 0$.

Special case 2: 2×2 *table.* Here we assume we have two treatments and a binary response, as in the data of Table 1 for example, so the data for each study take the form of two independent binomial distributions $f_1 \sim \mathrm{bin}(m_1, p_1)$ and $f_2 \sim \mathrm{bin}(m_2, p_2)$. Each

study gives an estimated log-odds ratio $\hat{\theta}$ in (6) with estimated variance $\hat{\sigma}^2$ in (7). The corresponding study parameters are

$$\theta = \log \frac{p_1(1-p_2)}{(1-p_1)p_2} \quad \text{and} \quad \sigma^2 = \gamma_1 + \gamma_2 \ , \tag{13}$$

where

$$\gamma_j = \frac{n}{m_j p_j (1-p_j)} \ \ ; \ j = 1, 2. \tag{14}$$

In this case, $\hat{\theta}$ and $\hat{\sigma}^2$ can be substantially correlated, and we show in Appendix A2 that for large n,

$$b = -\frac{d}{a} \ , \ d = \frac{\gamma_2^2(1-2p_2) - \gamma_1^2(1-2p_1)}{2(\gamma_1+\gamma_2)^2} \ . \tag{15}$$

With estimated within-study variances, $\tilde{\theta}$ is no longer an unbiased estimate of θ, but has a bias of order $O(N^{-1})$. This means that if N is large the bias in $\tilde{\theta}$ is small compared to its standard error which is of order $O(N^{-\frac{1}{2}})$. Similarly, it turns out that $\mathrm{E}(Q) = k-1+O(N^{-1})$. However, for tests and confidence intervals we need the quantities Z_1, Z_2, Z_3, and each of these has a bias of order $O(N^{-\frac{1}{2}})$. The size of the bias depends on the values of (a, b, d) in each study. Explicitly, we show in Appendix A3 that

$$\mathrm{E}(Z_j) = A_j N^{-\frac{1}{2}} + O(N^{-1}) \ \ ; \ j = 1, 2, 3 \ , \tag{16}$$

where

$$A_1 = \left(\frac{k}{s_{aa}}\right)^{\frac{1}{2}} \left\{ s_{ab} - \bar{d} + \frac{s_{a^2 d}}{k s_{aa}} \right\} \ , \tag{17}$$

$$A_2 = \left(\frac{k}{c_{aa} s_{aa}^3}\right)^{\frac{1}{2}} \times$$

$$\left[s_{aa}(c_{aa}\bar{b} - c_{ab}\bar{a}) - \frac{1}{k}\left\{ c_{ad}(2\bar{a}^2 - s_{aa}) - (k-2)s_{aa}\bar{a}\bar{d} - \frac{\bar{a} c_{(a-\bar{a})^2 d}}{c_{aa}}(2s_{aa} - \bar{a}^2) \right\} \right] \ , \tag{18}$$

and

$$A_3 = \left(\frac{k}{c_{aa}}\right)^{\frac{1}{2}} \left\{ c_{ab} - \left(1 - \frac{2}{k}\right)\bar{d} + \frac{c_{(a-\bar{a})^2 d}}{k c_{aa}} \right\} \ . \tag{19}$$

Note that the study-specific values (a, b, d) enter into the bias terms through the averages and average squares and products $\bar{a}$, s_{aa}, s_{ab} etc.

3 Bias corrections

We show in Appendix A3 that

$$\mathrm{Var}(Z_j) = 1 + O(N^{-1}) \ ; \ j = 1, 2, 3 \ . \tag{20}$$

Comparing this with (16) we see that for large N the shift in the distribution of Z_j is in the mean rather than the variance. This confirms the impression from Figure 2, which

compared an estimate of the density of Z_1 in the medical example with the density of the standard normal.

If we ignore terms of order $O(N^{-1})$ but allow for the bias in Z_1 of order $O(N^{-\frac{1}{2}})$, a normal approximation for Z_1 gives the adjusted confidence interval for θ with limits

$$\tilde{\theta} - (ks_{xx})^{-\frac{1}{2}}\{\hat{A}_1 N^{-\frac{1}{2}} \pm z_\alpha\} \ , \tag{21}$$

where $\hat{A}_1$ is an estimate of A_1 in (17) and z_α is the standard normal percentage point needed to achieve the desired coverage $(1-\alpha)$. The conventional confidence interval is just (21) with the bias term omitted. The analogous two-sided P-value for treatment effect is

$$P = 2\Phi\{-|(ks_{xx})^{\frac{1}{2}}\tilde{\theta} - \hat{A}_1 N^{-\frac{1}{2}}|\} \ , \tag{22}$$

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

$$2\Phi\{-|Z_2 - \hat{A}_2 N^{-\frac{1}{2}}|\} \ \text{ and } \ 2\Phi\{-|Z_3 - \hat{A}_3 N^{-\frac{1}{2}}|\} \ . \tag{23}$$

To implement these adjustments we have to estimate A_1, A_2 and A_3 using (17), (18) and (19). Again these depend on the nature of the studies being combined.

Special case 1: normal data. Here $b = d = 0$ for all studies, so all three quantities A_j are zero. No bias corrections are needed in this case. Of course the procedures are still biased, but the biases are of a lower order of magnitude.

Special case 2: 2×2 *tables.* From (12) and (13),

$$a = \sqrt{\frac{n}{N(\gamma_1 + \gamma_2)}} \ . \tag{24}$$

This, together with (15), shows that (a, b, d) 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 a common log-odds ratio. The constrained maximum likelihood estimates of p_1 and p_2 given that

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

are

$$p_1(\lambda) = \frac{f_1 + \lambda + 0.5}{m_1 + 1} \quad \text{and} \quad p_2(\lambda) = \frac{f_2 - \lambda + 0.5}{m_2 + 1} \ , \tag{25}$$

where λ is a Lagrange multiplier given by solving the quadratic equation

$$\Theta\{p_1(\lambda), p_2(\lambda)\} = \tilde{\theta} \ . \tag{26}$$

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 (26) has two real solutions and that it is the larger solution for λ which gives values of p_1 and p_2 in $[0, 1]$. Thus, to estimate (a, b, d) for each study, we find λ from (26), (p_1, p_2) from (25), (γ_1, γ_2) from (14) and hence (a, b, d) from (24) and (15).

Since the special case of 2×2 tables is the most commonly occurring example in medical applications, we now set out these 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 (6) and (7), and hence the usual radial plot co-ordinates (x, y) in (11) and the radial plot regression statistics of interest.

- Let $\eta = \exp(\tilde{\theta})$. If $\eta \neq 1$, calculate the Lagrange multipliers

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

  and hence each study's estimate of (p_1, p_2) in (25). If $\eta = 1$ (no treatment effect) then we simply take $\hat{p}_1 = \hat{p}_2 = (f_1 + f_2 + 1)/(m_1 + m_2 + 2)$. We can now calculate the estimates $(\hat{\gamma}_1, \hat{\gamma}_2)$ from (14) and hence the following three quantities for each study:

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

- Calculate the average values across the k studies of the quantities $(\hat{a}, \hat{b}, \hat{d}, \hat{a}^2, (\hat{a} - \bar{\hat{a}})^2)$ and the empirical mean squares and covariances needed to find $\hat{A}_1$, $\hat{A}_2$ or $\hat{A}_3$ from (17), (18) or (19). These give the bias corrections for the tests or confidence intervals specified in (21), (22) and (23).

To illustrate these calculations we return to the medical example mentioned in Section 1, with data set out in Table 1. First we fit the fixed effects model to the observed 2×2 tables by estimating (p_1, p_2) as explained above, and hence calculate estimates of (a, b, 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 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 usual 95% confidence interval

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

The bias estimate $\hat{A}_1 N^{-\frac{1}{2}}$ is $+0.879$ giving the corrected confidence interval

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

In this case $\tilde{\theta}$ is sufficiently accurate that the bias correction results in only a small shift in the interval, 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 (9) is $Z_2 = -2.844$ giving the conventional P-value as $P = 0.004$, highly significant evidence for publication bias. The bias estimate $\hat{A}_2 N^{-\frac{1}{2}}$ is -0.927, and so the bias adjusted P-value in (23) 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 also 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, following Copas and Malley (2008), 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 (10) 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 $\hat{A}_3 N^{-\frac{1}{2}}$ is -0.0818, adjusting the P-value in (23) to $P = 0.106$. In this case the bias correction to Z_3 is unimportant, but allowing for publication bias has been catastrophic as far as the conclusion of this particular meta analysis is concerned.

4 Simulation study

In this section we give a very brief summary of some simulation results exploring the accuracy of our approximations. We only report results here for the special case of 2×2 tables, and for perhaps the two most common tasks in meta analysis, confidence intervals and P-values for θ, and the intercept (Egger) test for publication bias. Clearly it is impossible to cover all possibilities: we illustrate our results by looking at the distributions of Z_1 and Z_2 in 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 3. The simulations are repeated $100,000$ times.

Figure 3 looks at Z_1 with $\theta = \log(0.2)$ and shows kernel density estimates of the distribution of Z_1, and of the bias corrected version $Z_1 - \hat{A}_1 N^{-\frac{1}{2}}$. The distribution of $Z_1 - \hat{A}_1 N^{-\frac{1}{2}}$ 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 $\mathrm{E}(Z_1)$ have been very effective at removing the bias.

Figure 4 takes $\theta = \log(0.5)$ and shows kernel density estimates of the distribution of Z_2 and of the bias corrected version $Z_2 - \hat{A}_2 N^{-\frac{1}{2}}$. Here there is an even larger shift, this time to the left, but again the distribution of $Z_2 - \hat{A}_2 N^{-\frac{1}{2}}$ 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 particulary 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). In all rows of the table, bias is indicated by an imbalance between the left and right tail error rates.

Radial plot methods apply to meta analysis problems rather generally, but for the specific case of 2×2 tables other methods are also available, notably the Mantel-Haenszel estimate (Mantel and Haenszel, 1959; Mantel, 1963). This estimate works directly on odds ratios without the logarithmic transformation, and so is not a function of the radial plot as defined here. For completeness we have included in Table 2 the analogous results for the Mantel-Haenszel method, using the analogue of Z_1 for the logarithm of the Mantel-Haenszel estimate and the variance estimate of Robins *et al.* (1986). For the intercept test (Egger test) based on Z_2, we can also compare our results with the recently proposed test of Harbord *et al.* (2006). Again their test is specific to the 2×2 case and is not a function of the radial plot. In Table 2 we include error rates for the Harbord *et al.* test statistic in place of Z_2.

θ	statistic	left tail %	right tail %	significance level %
$\log(0.67)$	Z_1	3.71	6.21	9.91
	Mantel-Haenszel	4.06	5.96	10.02
	$Z_1 - \hat{A}_1 N^{-\frac{1}{2}}$	4.98	5.06	10.03
	Z_2	20.62	0.72	21.35
	Harbord *et al.*	4.85	5.10	9.95
	$Z_2 - \hat{A}_2 N^{-\frac{1}{2}}$	4.99	5.20	10.19
$\log(0.5)$	Z_1	3.19	7.11	10.30
	Mantel-Haenszel	3.59	6.59	10.18
	$Z_1 - \hat{A}_1 N^{-\frac{1}{2}}$	5.01	4.89	9.90
	Z_2	38.22	0.12	38.34
	Harbord *et al.*	5.04	5.06	10.10
	$Z_2 - \hat{A}_2 N^{-\frac{1}{2}}$	4.48	5.14	9.62
$\log(0.2)$	Z_1	1.64	11.85	13.49
	Mantel-Haenszel	2.30	9.91	12.21
	$Z_1 - \hat{A}_1 N^{-\frac{1}{2}}$	4.97	5.18	10.14
	Z_2	88.12	0	88.12
	Harbord *et al.*	7.63	2.50	10.13
	$Z_2 - \hat{A}_2 N^{-\frac{1}{2}}$	1.88	6.89	8.77

Table 2. Error rates for nominal 10% (two-sided) and 5% (one-sided) confidence intervals and tests.

These simulations, plus more extensive results covering other configurations, suggest

some tentative conclusions about how the biases of Z_1, Z_2 and Z_3 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 part of Table 2 where the Z_2 test is grossly misleading: at this setting there are likely to be several zeros in the data).

Table 2 suggests that the Mantel-Haenszel method suffers a bias rather similar to that of Z_1, suggesting that a bias correction to this estimate would also be useful. Table 2 also shows (at least for the cases considered) that the test of Harbord *et al.* (2006) is effective in reducing the bias of the Egger test. At more extreme values of θ, seen in the last part of Table 2, there is a tendency for Harbord *et al.* to under-correct, and for $Z_2 - \hat{A}_2 N^{-\frac{1}{2}}$ to over-correct, for this bias, although the (two-tail) significance levels are quite similar (and dramatically better than the significance level of Z_2).

5 Conclusions and comments

1. The variance-weighted estimate $\tilde{\theta}$ in (2), widely used in meta analysis, can be thought of as the slope of a linear regression through the origin of the radial plot (plot of standardized treatment effects against one over their standard errors). Tests and confidence intervals for θ based on $\tilde{\theta}$, as well as a number of other meta analysis tasks, can similarly be reduced to regression-like calculations on the radial plot. Under the naive assumption that within-study variances are known, statistical properties of methods of meta analysis then follow immediately from those of standard linear regression. However, using estimated within-study variances means that the radial plot no longer satisfies the usual assumptions of linear regression, leading to a bias in $\tilde{\theta}$ and related inferences. We have suggested a rather general

setting for investigating and correcting for these biases. At least for the cases discussed above, these biases can be quite substantial. Provided $|\theta|$, the size of the treatment effect, is not too large, the bias corrections proposed here seem effective and will be useful in practice.

2. The bias corrections in Section 3 depend on the quantities (a, b, d) for each study, and hence on the statistical properties of the studies being combined. We have evaluated these explicitly for the case of normally distributed observations and for the case of 2×2 tables. However, meta analysis is used much more widely, for example the case of estimated log hazard ratios in survival studies would also be of interest. Insight into the statistical properties of $(\hat{\theta}, \hat{\sigma}^2)$ is needed for estimating (a, b, d) and so is a pre-requisite for estimating bias corrections for the resulting meta analysis.

3. 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 often uses the estimated residual mean square as in standard regression analysis. Although less efficient than Z_2 under the fixed effects model, using the observed residual mean square has the advantage of being more robust to heterogeneity between the studies.

4. For completeness we have also included the Mantel-Haenszel and Harbord *et al.* methods in Table 2 since these are also available in the special case of 2×2 tables. There is a large literature on the Mantel-Haenszel estimate and how it compares with the variance-weighted approach, see Sutton *et al.* (2000, section 4.3.5) for a summary. The consensus in most of this literature is that Mantel-Haenszel is better if there is a large number of small studies, but $\tilde{\theta}$ is better for a small number of large studies. It is unclear how these conclusions would be affected if bias corrections were introduced. The test proposed by Harbord *et al.* (2006) is just one of several other recently proposed improvements to the Egger test (Macaskill *et al.*, 2001; Peters *et al.*, 2006; Rücker *et al.*, 2008; Schwarzer *et al.*, 2007). These may be effective (and simpler) alternatives to using the bias correction for Z_2, but their comparative properties have yet to be fully evaluated.

5. We have assumed the fixed effects model throughout. The usual practice is to test this assumption by calculating Q in (8) 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, section 5.2). 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 $\mathrm{E}(Z_1)$ 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 2 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. The literature shows that allowing for the uncertainty in $\hat{\tau}^2$, especially with the truncation usually used to avoid negative variance estimates, is difficult enough without adding the complication of uncertainty in the $\hat{\sigma}_i^2$'s.

6. The bias formulae for the case of 2×2 tables 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 of these hold (balanced trials with no treatment effect), the biases are all zero since $d = 0$ in (15). In particular, both Z_1 and Z_3 provide approximately unbiased tests for the significance of the overall treatment effect in the important case of balanced clinical trials, when no bias corrections are needed.

Acknowledgement

We are grateful to referees for their helpful comments. One of us (CL-C) was supported by CONACYT (Mexico) grant no. 160859.

References

Copas, J. B. and Malley, P. F. (2008). A robust P-value for treatment effect in meta analysis with publication bias. *Statist. in Med.*, **27**, 4267-4278.

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.

Mantel, N. (1963). Chi-squared tests with one degree of freedom: extension of the Mantel-Haenszel procedure. *J. Am. Statist. Assoc.*, **58**, 690-700.

Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J. Nat. Cancer Inst.*, **22**, 719-748.

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.

Robins, J., Breslow, N and Greenland, S. (1986). Estimators of the Mantel-Haenszel variance consistent in both sparse and large-stratum limiting models. *Biometrics*, **42**, 311-323.

Rücker, G., Carpenter, J. R. and Schwarzer, G. (2008). Arcsin tests for publication bias. *Statist. in Med.*, **27**, 746-763.

Schwarzer, G., Antes, G. and Schumacher, 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.

Schwarzer, G., Antes, G. and Schumacher, 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., Sheldon, T. A. and Song, F. (2000). *Methods for meta-analysis in medical research.* Chichester: Wiley.

Appendix

Here is an outline of the derivation of some of the formulae quoted in Section 2 and 3:

A1 : *Approximations 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 $\mathrm{Var}(\hat{\theta}) = \sigma^2/n + O(n^{-2})$, and that $\mathrm{Var}(\hat{\sigma}^2) = O(n^{-1})$. These are standard properties of maximum likelihood estimates; in practice they will hold for any 'sensible' estimates of θ and σ^2.

Now define

$$u = x - \frac{\sqrt{n}}{\sigma} = \sqrt{n}\left(\frac{1}{\hat{\sigma}} - \frac{1}{\sigma}\right), \quad v = y - \frac{\theta\sqrt{n}}{\sigma} = \sqrt{n}\left(\frac{\hat{\theta}}{\hat{\sigma}} - \frac{\theta}{\sigma}\right), \quad w = v - \theta u = \frac{(\hat{\theta} - \theta)\sqrt{n}}{\hat{\sigma}}. \tag{27}$$

Then

$$b = N^{\frac{1}{2}}\mathrm{E}(w) \quad \text{and} \quad d = -\mathrm{E}(uw) \ ,$$

both of which are of order $O(1)$ as $N \to \infty$. Also as a consequence of the above assumptions we have

$$\mathrm{E}(u) = O(n^{-\frac{1}{2}}) \ , \quad \mathrm{Var}(w) = 1 + O(n^{-1}) \ \text{ and } \ \mathrm{Cov}(w, uw) = O(n^{-\frac{1}{2}}). \tag{28}$$

A2 : *Special case 2:* 2×2 *table*

First consider the case of a single binomial distribution, with observed frequency $f \sim \mathrm{bin}(m, p)$, say. Let z be the asymptotic (large m) standard normal deviate corresponding to f, that is

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

Then after some tedious but straightforward algebra we find

$$\log\frac{f+\frac{1}{2}}{m-f+\frac{1}{2}} = \log\frac{p}{1-p} + \gamma^{\frac{1}{2}} z m^{-\frac{1}{2}} + \frac{1}{2}\gamma(1-2p)(1-z^2)m^{-1} + O_p(m^{-3/2}), \quad (29)$$

$$\frac{m+1}{(f+\frac{1}{2})(m-f+\frac{1}{2})} = \gamma m^{-1} - \gamma^{3/2}(1-2p)zm^{-3/2} + O_p(m^{-2}). \quad (30)$$

Now extend 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 $n = m_1 + m_2$ and $(\hat{\theta}, \hat{\sigma}^2)$ defined in (6) and (7). We use (29) and (30) 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. With (14) this leads to

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

and

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

which in turn lead to (15).

A3 : *Approximations for radial plot statistics*

From (27) and (12) we get

$$x = \frac{\sqrt{n}}{\hat{\sigma}} = aN^{\frac{1}{2}} + u \quad \text{and} \quad y = \frac{\hat{\theta}\sqrt{n}}{\hat{\sigma}} = a\theta N^{\frac{1}{2}} + v\ .$$

Hence

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

These lead to

$$\tilde{\theta} = \theta + \frac{s_{aw}}{s_{aa}}N^{-\frac{1}{2}} + O_p(N^{-1}) \quad (31)$$

$$Z_1 = \left(\frac{k}{s_{aa}}\right)^{\frac{1}{2}}\left\{s_{aw} + \left(s_{uw} - \frac{s_{aw}s_{au}}{s_{aa}}\right)N^{-\frac{1}{2}}\right\} + O_p(N^{-1}) \quad (32)$$

$$Q = k\left\{s_{ww} - \frac{s_{wa}^2}{s_{aa}} - \frac{2s_{wa}}{s_{aa}^2}(s_{aa}s_{uw} - s_{au}s_{aw})N^{-\frac{1}{2}}\right\} + O_p(N^{-1})\ . \quad (33)$$

The expectations of the random quantities appearing in (31) to (33) are

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

and

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

These lead to $\mathrm{E}(\tilde{\theta}) = \theta + O(N^{-1})$, $\mathrm{E}(Q) = k - 1 + O(N^{-1})$, and (17).

The analogous expressions for c_{xx} and c_{xy} are essentially the same but with the obvious replacement of terms like s_{aa} with c_{aa}. Reworking these calculations for Z_3 leads to (19). For Z_2 we find

$$\bar{y} - \breve{\theta}\bar{x} = \bar{w} - \frac{c_{aw}\bar{a}}{c_{aa}} - \left\{ \frac{c_{aw}\bar{u}}{c_{aa}} + \frac{\bar{a}(c_{uw}c_{aa} - 2c_{aw}c_{au})}{c_{aa}^2} \right\} N^{-\frac{1}{2}} + O_p(N^{-1}) \ ,$$

which leads to (18).

Finally, from (32) we have

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

where B is the factor multiplying $N^{-\frac{1}{2}}$ in (32). But B can be written as a linear combination of products of the form $u_i w_j$, and so from (28) we get

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

Equation (20) follows for $j = 1$, and similarly for $j = 2, 3$.

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) confidence intervals and 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.

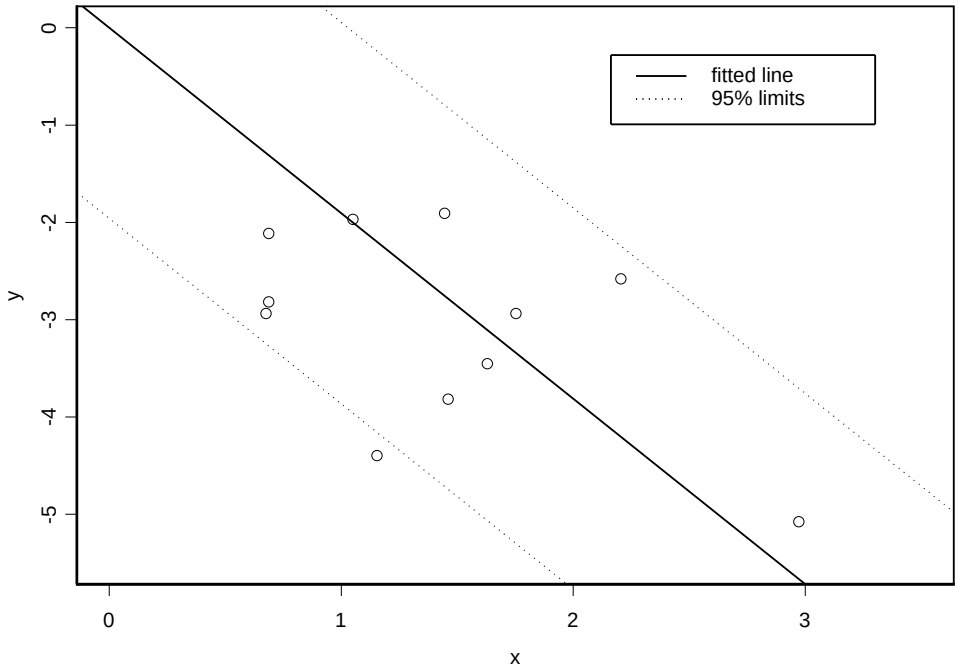

Figure 1: Radial plot for iron supplementation meta analysis.

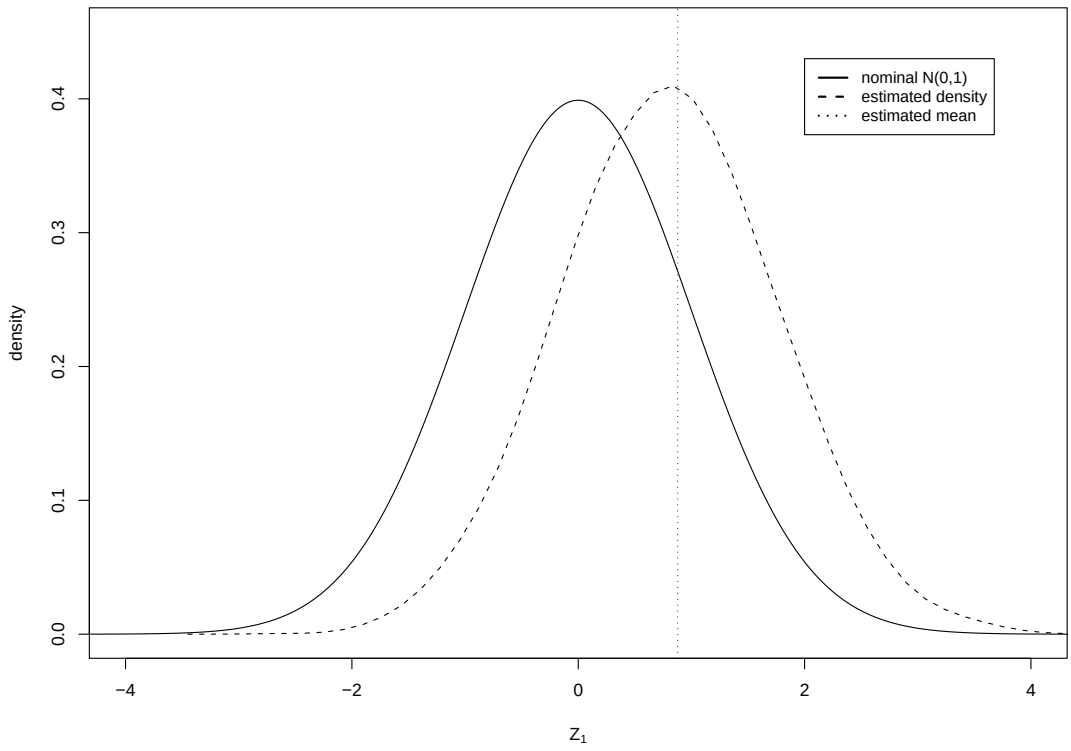

Figure 2: Distribution of Z_1 for iron supplementation meta analysis.

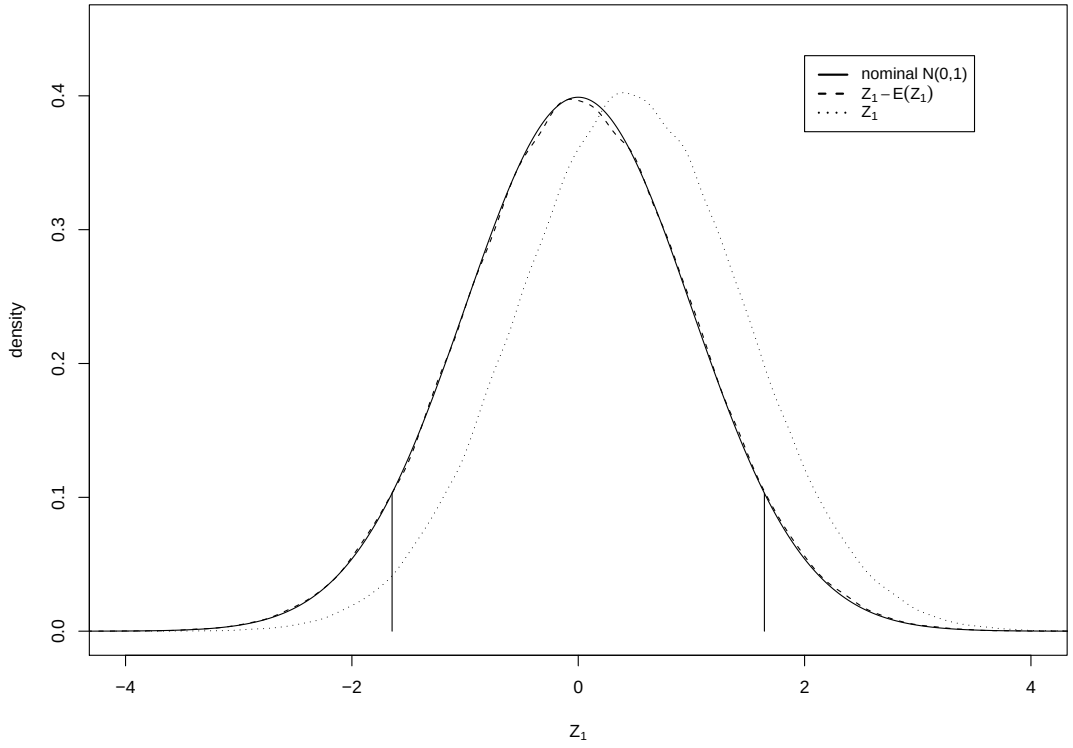

Figure 3: Simulation study: distribution of Z_1.

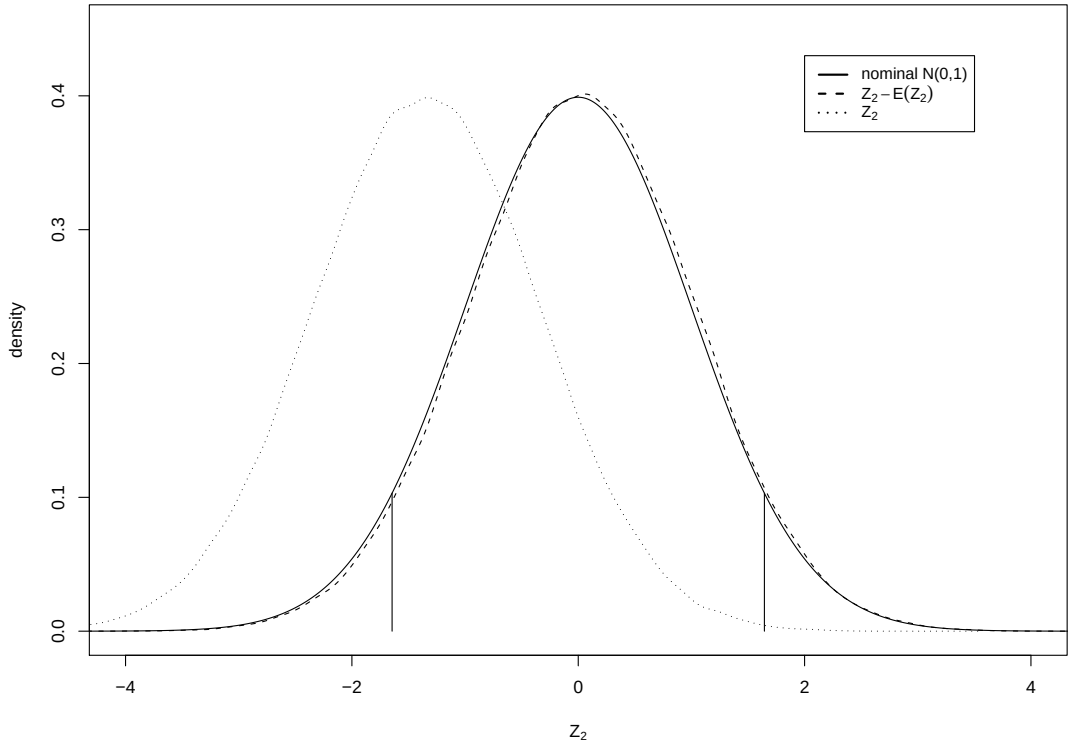

Figure 4: Simulation study: distribution of Z_2.