STATISTICS IN TRANSITION new series, December 2021
Vol. 22, No. 4, pp. 1–17, DOI 10.21307/stattrans-2021-035
Received – 04.05.2021; accepted – 06.09.2021

Unreported standard errors in meta-analysis

Nicholas T. Longford[1]

ABSTRACT

A study that would otherwise be eligible is commonly excluded from a meta-analysis when the standard error of its treatment-effect estimator, or the estimate of the variance of the outcomes, is not reported and cannot be recovered from the available information. This is wasteful when the estimate of the treatment effect is reported. We assess the loss of information caused by this practice and explore methods of imputation for the missing variance. The methods are illustrated on two sets of examples, one constructed specifically for illustration and another based on a published systematic review.

Key words: empirical Bayes, imputation, meta-analysis, missing value, sensitivity analysis.

1. Introduction

In a typical meta-analysis for comparing two treatments, A and B, there are H studies and for each study i we have an estimate $\hat\theta_i$ of the treatment effect θ_i, an estimate $\hat\sigma_i^2$ of the variance σ_i^2 of the outcomes and the within-treatment sample sizes $n_{i\mathrm{A}}$ and $n_{i\mathrm{B}}$, from which the standard error of $\hat\theta_i$, denoted by τ_i, can be easily estimated. For example, when the subjects in study i are assigned to the treatments completely at random subject to fixed sample sizes $n_{i\mathrm{A}}$ and $n_{i\mathrm{B}}$, we have $\tau_i^2 = \sigma_i^2(1/n_{i\mathrm{A}} + 1/n_{i\mathrm{B}})$, and τ_i^2 is estimated by $\hat\tau_i^2 = \hat\sigma_i^2(1/n_{i\mathrm{A}} + 1/n_{i\mathrm{B}})$. We assume that the estimators $\hat\theta_i$ and $\hat\sigma_i^2$ are unbiased for the respective targets θ_i and σ_i^2, and that the variances $\sigma_{i\mathrm{A}}^2$ and $\sigma_{i\mathrm{B}}^2$ within the two treatment groups coincide with σ_i^2. The development presented here can easily be adapted for heteroscedasticity because the key parameter we work with is the standard error τ_i and its estimate. Note that $\hat\tau_i$ is not unbiased for τ_i, and neither is $1/\hat\tau_i^2$ for $1/\tau_i^2$, even when $\hat\tau_i^2$ is unbiased for τ_i^2; see Longford (2010 and 2015) for a discussion of this issue in a wider context.

For background to meta-analysis we refer to Rice, Higgins and Lumley (2018) and references therein. Of historical importance is Glass (1976), credited with coining the term, and Hedges and Olkin (1985), the first comprehensive account of statistical methods for meta-analysis. Nowadays, meta-analysis is applied widely, in social and medical sciences in particular, to pool information across studies in which identical or closely related parameters are estimated.

Systematic reviews are a formalised approach to identifying studies suitable for meta-analysis and related purposes; see Haidich (2010) for an introduction. The CONSORT statement (Begg *et al.*, 1996) and the STROBE initiative (von Elm *et al.*, 2008) formulate guidelines and standards for the conduct and presentation of such reviews and for reporting

[1]School of Public Health, Imperial College London, UK. E-mail: sntlnick@sntl.co.uk.
ORCID: https://orcid.org/0000-0003-4129-9726.

single case studies in a manner conducive to their use in future systematic reviews. They are widely adopted today.

We are concerned with the setting in which an estimate of the sampling variance τ_i^2 is not available for one or a few studies. We deal with the case of a single study for which $\hat{\tau}_i^2$ is not available, but the proposed methods and conclusions carry over to meta-analysis in which several studies have this deficiency. Our focus is on meta-analysis with only a few studies, to which even a single study may contribute with a relatively large amount of information, so we can ill-afford to discard it. We assume that the estimates $\hat{\theta}_i$ and the sample sizes $n_{i\mathrm{A}}$ and $n_{i\mathrm{B}}$ are available for all studies.

There are two generic methods for dealing with missing values in an analysis. By list-wise deletion, we apply the planned analysis to the units (studies in a meta-analysis) for which we have complete information. This is wasteful because we discard some studies even though we have their estimates $\hat{\theta}_i$, and sometimes also the sample sizes n_i and other details. By imputation, we substitute a value for each missing data item. However well we may estimate the missing values $\hat{\sigma}_i^2$ (or $\hat{\tau}_i^2$), we overstate the precision of the estimator $\hat{\theta}$ of the overall treatment effect θ because by treating the imputed values $\tilde{\sigma}_i^2$ (or $\tilde{\tau}_i^2$) on par with the corresponding estimates we pretend to have more information than was in fact collected. Multiple imputation (Rubin, 2004) addresses this deficiency in a principled way, although it entails some complexities in our context.

Various forms of sensitivity analysis can hone in on the range of plausible values of the complete-data estimator of the average treatment effect. For outcomes with values in a finite range, imputation of extreme values is an obvious starting point. For an improvement of this method, see Gamble and Hollis (2005). Publication bias is another issue related to missing values. It concerns studies that were conducted but their results were not published. For a landmark contribution to this topic, see Duval and Tweedie (2000). Rothstein, Sutton and Borenstein (2005) is an authoritative edited volume dedicated to this subject. See Lin and Chu (2018) for a recent contribution.

Our problem relates to a study published with incomplete information. On the one hand, we want to rescue such a study for the meta-analysis by using all the available data; on the other hand, we want to reflect in the statements we make the loss due to the incompleteness. In brief, we want to be 'honest' in our inferential statements.

We explore two general approaches, modelling and sensitivity analysis. In Section 3, we specify an empirical Bayes model for the variances σ_i^2 and impute a random draw from the approximated conditional distribution of the missing variance σ_{H+1}^2. This imputation is replicated (independently repeated) several times, to generate a set of plausible completions of the dataset. We assume that the study-specific treatment effects coincide; $\theta_i = \theta$ for all studies i. In Section 3.2 we discuss random-effects meta-analysis (DerSimonian and Laird, 1986), in which this assumption is relaxed and the treatment effects θ_i are a random sample from an unknown distribution.

In Section 4, we apply a method motivated by sensitivity analysis, in which we consider a plausible range of values of σ_i^2, or τ_i^2, and evaluate the corresponding estimates of the overall effect θ and standard errors of $\hat{\theta}$. Section 5 applies the methods to a meta-analysis with complete information, in which the standard error of one study is masked. Section 6 discusses some peripheral issues; they include elicitation of the information about the

Table 1: Examples of sets of five studies included in a meta-analysis, with the sampling variance estimate $\hat{\tau}_i^2$ not available for one study.

	Case A			Case B			Case C		
Study (i)	$\hat{\theta}_i$	$\hat{\tau}_i^2$	n_i	$\hat{\theta}_i$	$\hat{\tau}_i^2$	n_i	$\hat{\theta}_i$	$\hat{\tau}_i^2$	n_i
1	0.467	0.260	80	0.617	0.260	80	0.567	0.260	80
2	0.082	0.365	46	0.232	0.365	46	0.182	0.365	46
3	0.384	0.229	102	0.534	0.229	102	0.484	0.229	102
4	0.163	0.282	66	0.313	0.282	66	0.263	0.282	66
5	0.691	?	92	0.691	?	92	0.621	?	92

missing value(s) and exploiting the information about the mean-variance relationship of the outcomes.

Table 1 presents three examples, A, B and C, of study results for meta-analysis, each with $H+1=5$ studies, on which we illustrate the methods we develop. In each example, all five studies have two treatment arms, with equal variances and equal sample sizes within the arms of each study; $\sigma_{i\mathrm{A}}^2 = \sigma_{i\mathrm{B}}^2 = \sigma_i^2$ and $n_{i\mathrm{A}} = n_{i\mathrm{B}} = \frac{1}{2}n_i$, $i = 1,\ldots,5$. The quintets of sample sizes n_i and the quartets of estimates of τ_i^2 are the same across the three cases, only the sets of estimates differ.

By back-calculating the within-treatment variance estimates we can check that the variances are very likely to differ; the estimates are in the range 8.4 – 11.7. Study 5, with $\hat{\tau}_5^2$ not available, has an unexceptional sample size. In each case A – C, we consider the plausible range $(0.17, 0.28)$ for $\hat{\tau}_5^2$. That is, we rule out the possibility that τ_5^2 may be smaller than 0.17 or larger than 0.28. This choice is informed by the sample size and the variances in the other studies. Some leeway at either limit of their range is allowed since the (unknown) variance may be larger or smaller than the four recorded variances. In practice, expert opinion may provide some additional input.

2. Information gained by using imputation

Suppose we have H studies with complete information and another study, $H+1$, with the value of $\hat{\tau}_{H+1}^2$ (or $\hat{\sigma}_{H+1}^2$) missing. The treatment effect common to the H studies is estimated by

$$\hat{\theta}_- = \frac{w_1\,\hat{\theta}_1 + \cdots + w_H\,\hat{\theta}_H}{W_H},$$

where $w_i = 1/\hat{\tau}_i^2$ and $W_H = w_1 + \cdots + w_H$. Ignoring the uncertainty about the weights w_i, that is, about the variances τ_i^2, leads to the expression $\mathrm{var}(\hat{\theta}_-) = 1/W_H$. This confirms that *information*, defined as the reciprocal of the sampling variance, is additive. Specifically, the information about θ contained in study i is w_i, in the collection of H studies it is W_H and, if τ_{H+1}^2 were available, it would be $W_H + w_{H+1}$ in the $H+1$ studies.

If we had complete information about study $H+1$, we would evaluate the version of the

estimator $\hat{\theta}_-$ for $H+1$ studies, that is,

$$\hat{\theta}_+ = \frac{W_H\,\hat{\theta}_- + w_{H+1}\,\hat{\theta}_{H+1}}{W_H + w_{H+1}}\,.$$

If w_{H+1} were known the variance of the estimator of θ would be reduced by

$$\frac{1}{W_H} - \frac{1}{W_H + w_{H+1}} = \frac{w_{H+1}}{W_H\,(W_H + w_{H+1})}\,,$$

or by $100 w_{H+1}/(W_H + w_{H+1})\%$. As w_{H+1} is not known, the potential for reduction is smaller. When w_{H+1} is not known, but a plausible range for it is defined, then we can find the plausible range of this percentage. In the cases in Table 1, this range is $(19.8, 28.9)\%$. The plausible reduction of the standard error is in the range $(9.5, 13.5)\%$. Thus, a lot is at stake; the sampling variance could be reduced by as much as 29%, but the uncertainty about the magnitude of this stake, about 9%, is not trivial either.

3. Empirical Bayes model for σ_i^2

Imputation for a variance estimate is based on an estimate of the distribution underlying the variances of the studies. We assume that this distribution is inverse gamma, and estimate its parameters. First we derive the marginal distribution of the estimator $\hat{\sigma}_i^2$ of the within-treatment group variance σ_i^2 in study $i = 1, \ldots, H$.

We assume that, conditionally on the variance σ_i^2, $k_i\,\hat{\sigma}_i^2/\sigma_i^2$ has χ^2 distribution with k_i degrees of freedom. Thus, the conditional density of $\hat{\sigma}_i^2$, given its estimand σ_i^2, is

$$f(x) = \frac{1}{\Gamma\left(\frac{1}{2}k_i\right)} \left(\frac{k_i}{2\sigma_i^2}\right)^{\frac{1}{2}k_i} x^{\frac{1}{2}k_i - 1} \exp\left(-\frac{k_i x}{2\sigma_i^2}\right),$$

where Γ is the gamma function. Further, we assume that the variances σ_i^2 are a random sample from the inverse gamma distribution with parameters α and γ:

$$g(y) = \frac{1}{\Gamma(\gamma)}\,\alpha^\gamma \left(\frac{1}{y}\right)^{\gamma+1} \exp\left(-\frac{\alpha}{y}\right).$$

The marginal density of $\hat{\sigma}_i^2$ is obtained by integration of the joint density of $\hat{\sigma}_i^2$ and σ_i^2:

$$\begin{aligned} & C x^{\frac{1}{2}k_i - 1} \int_0^{+\infty} \left(\frac{1}{y}\right)^{\frac{1}{2}k_i + \gamma + 1} \exp\left\{-\frac{1}{y}\left(\alpha + \frac{1}{2}k_i x\right)\right\} \mathrm{d}y \\ = \;& C\Gamma\left(\frac{k_i}{2} + \gamma\right) \frac{x^{\frac{1}{2}k_i - 1}}{\left(\alpha + \frac{1}{2}k_i x\right)^{\frac{1}{2}k_i + \gamma}}\,, \end{aligned}$$

where C is the standardising constant, for which the expression is a density. We approximate the concluding expression by an inverse gamma density using the relation $(1 + c/k)^k \doteq \mathrm{e}^c$,

precise for sufficiently large k. For the denominator, we have

$$\begin{aligned}\left(\alpha+\frac{k_i x}{2}\right)^{\frac{1}{2}k_i+\gamma} &= \left(\frac{k_i x}{2}\right)^{\frac{1}{2}k_i+\gamma}\left(1+\frac{2\alpha}{k_i x}\right)^{\frac{1}{2}k_i+\gamma} \\ &\doteq \left(\frac{k_i x}{2}\right)^{\frac{1}{2}k_i+\gamma}\exp\left\{\frac{2\alpha}{k_i x}\left(\frac{1}{2}k_i+\gamma\right)\right\}.\end{aligned}$$

Hence the approximation to the marginal density of $\hat{\sigma}_i^2$ by an inverse gamma density,

$$\frac{1}{\Gamma(\gamma)}\left\{\frac{\alpha}{k_i}(k_i+2\gamma)\right\}^{\gamma}\left(\frac{1}{x}\right)^{\gamma+1}\exp\left\{-\frac{\alpha}{k_i x}(k_i+2\gamma)\right\},$$

where the first two factors standardise the expression to be a density. The expectation of this distribution is $\mu=\alpha(1+2\gamma/k_i)/(\gamma-1)$, assuming that $\gamma>1$, and its variance is $\mu^2/(\gamma-2)$, assuming that $\gamma>2$. The parameters α and γ of this density are estimated by maximising the loglikelihood

$$l=-H\log\{\Gamma(\gamma)\}+H\gamma\log(\alpha)+\gamma\sum_{i=1}^{H}\log\left(\frac{k_i+2\gamma}{k_i}\right)-(\gamma+1)\sum_{i=1}^{H}\log\left(\hat{\sigma}_i^2\right)-\alpha\sum_{i=1}^{H}\frac{k_i+2\gamma}{k_i\hat{\sigma}_i^2}.$$

We apply the Newton-Raphson algorithm. The score functions for l are

$$\begin{aligned}\frac{\partial l}{\partial\alpha} &= \frac{H\gamma}{\alpha}-\sum_{i=1}^{H}\frac{k_i+2\gamma}{k_i\hat{\sigma}_i^2} \\ \frac{\partial l}{\partial\gamma} &= -H\Gamma'(\gamma)+H\log(\alpha)+\sum_{i=1}^{H}\log\left(\frac{k_i+2\gamma}{k_i}\right)+2\gamma\sum_{i=1}^{H}\frac{1}{k_i+2\gamma}-\sum_{i=1}^{H}\log\left(\hat{\sigma}_i^2\right) \\ &\quad -2\alpha\sum_{i=1}^{H}\frac{1}{k_i\hat{\sigma}_i^2},\end{aligned}$$

where Γ' is the digamma function, the derivative of $\log(\Gamma)$. The elements of the Hessian matrix are

$$\begin{aligned}-\frac{\partial^2 l}{\partial\alpha^2} &= \frac{H\gamma}{\alpha^2} \\ -\frac{\partial^2 l}{\partial\alpha\,\partial\gamma} &= -\frac{H}{\alpha}+2\sum_{i=1}^{H}\frac{1}{k_i\hat{\sigma}_i^2} \\ -\frac{\partial^2 l}{\partial\gamma^2} &= H\Gamma''(\gamma)-4\sum_{i=1}^{H}\frac{1}{k_i+2\gamma}+4\gamma\sum_{i=1}^{H}\frac{1}{(k_i+2\gamma)^2},\end{aligned}$$

where Γ'' denotes the trigamma function, the derivative of the digamma function. The Newton-Raphson algorithm converges very fast, as judged by any reasonable criterion for convergence. An initial solution has to be provided; this is difficult to automate because the loglikelihood is not concave throughout the parameter space.

The expression for l implies that the sufficient statistics for α and γ are the average (or total) of $\log(\hat\sigma_i^2)$ and, assuming that $\gamma \ll \frac{1}{2}k_i$ for all i, the average (or total) of $1/\hat\sigma_i^2$. The 'weights' k_i are therefore not as important for summarising the variances σ_i^2 as they are for the treatment effect θ. We confirm this on an example in Section 5.

We derive a non-iterative estimator of (α,γ) that can be used as an alternative, or as an initial solution for the Newton-Raphson algorithm. Given $\hat\gamma$, $\partial l/\partial\alpha$ implies a simple expression for $\hat\alpha$;

$$\hat\alpha = \frac{H\hat\gamma}{\sum_{i=1}^{H}\dfrac{k_i+2\hat\gamma}{k_i\hat\sigma_i^2}}, \tag{1}$$

which is well approximated by $H\hat\gamma/(1/\hat\sigma_1^2+\cdots+1/\hat\sigma_H^2)$ when $\hat\gamma \ll k_i$ for all i.

The posterior distribution of σ_i^2 is inverse gamma with expectation $\mathrm{E}(\sigma_i^2\,|\,\hat\sigma_i^2)=c$ and variance $\mathrm{var}(\sigma_i^2\,|\,\hat\sigma_i^2)=c^2/(\gamma-2)$, where $c=\alpha(k+2\gamma)/\{k(\gamma-1)\}$. Denote these moments by E and V, respectively. The ratio E^2/V is equal to $\gamma-2$ for all k_i. This motivates the moment-matching estimator $\hat\gamma=2+\hat E^2/\hat V$, based on the naïve estimators of E and V. For α we do not have a moment-matching estimator, but we can use the estimator given by equation (1), without the assumption that $\gamma \ll k_i$. Problems with maximum likelihood are sometimes encountered with small-scale data or large values of $\log\{\Gamma(\hat\gamma)\}$. We have not come across any, but this non-iterative method can be regarded as a back-up for such an eventuality.

3.1. Imputation

With maximum likelihood estimators $\hat\alpha$ and $\hat\gamma$, we have several options for imputation for an unknown variance. The simplest is to impute the naïve estimator of the expectation of the fitted distribution, $\hat c=\hat\alpha(k+2\hat\gamma)/\{k(\hat\gamma-1)\}$. This quantity depends on the degrees of freedom k, although only weakly when $k \gg 2\gamma$, when $(k+2\hat\gamma)/k \doteq 1$. Next, we could use for imputation values generated by a draw from the fitted (inverse gamma) distribution. And finally, the uncertainty about α and γ could be reflected by drawing first a plausible pair $(\tilde\alpha,\tilde\gamma)$ from the fitted distribution for (α,γ) and then drawing a value $\tilde\sigma_i^2$ from the plausible distribution given by $(\tilde\alpha,\tilde\gamma)$. Some approximation cannot be avoided in this process because the joint distribution of $(\hat\alpha,\hat\gamma)$ is known only asymptotically and is estimated by using estimates for the unknown parameters. Bayesian counterparts of these procedures can be implemented; $(\tilde\alpha,\tilde\gamma)$ is drawn from the joint posterior distribution of (α,γ). They also entail some approximation; the paucity of information about α and γ is unavoidable, especially if we have no means of faithfully representing the prior information about them and, indeed, when our prior information is scant. Care has to be exercised also in the choice of a flat prior to represent the absence of any such information.

The maximum likelihood estimators of α and γ, based on studies $1-4$, have very large sampling variances and the two estimators are highly correlated. When maximum likelihood (or any other method) is fitted to a small number of studies the process of using plausible values entails a lot variation.

3.2. Meta-analysis with random effects

We have assumed that the studies have a common expectation θ. It may be more appropriate to assume that the study-specific treatment effects θ_i are a random sample from a distribution with expectation θ and variance $\omega \geq 0$. If this variance were known, the optimal estimator of θ based on the first H studies would be

$$\hat{\theta} = \frac{1}{W_H(\omega)} \sum_{i=1}^{H} w_H(\omega)\, \hat{\theta}_i\,,$$

where $w_i(\omega) = 1/(\omega + \hat{\tau}_i^2)$ and $W_H(\omega) = w_1(\omega) + \cdots + w_H(\omega)$; see DerSimonian and Laird (1986). A profound difficulty in using or adapting this estimator is that ω is not known and, when H is small, is estimated with very low precision. Even if all H studies were very large, so that there would be very little uncertainty about each θ_i, $i = 1,\ldots,H$, ω could be estimated with only $H-1$ degrees of freedom. When the studies have moderate sample sizes and there is appreciable uncertainty about each $\hat{\theta}_i$, the uncertainty about ω is even greater. This is difficult to reflect in the estimation of $\mathrm{var}(\hat{\theta})$, but it is obvious that the conventional estimator $\widehat{\mathrm{var}}(\hat{\theta}) = 1/W_H(\hat{\omega})$ has a negative bias. In fact, even with the assumption of a common treatment effect, $\omega = 0$, the estimator $\widehat{\mathrm{var}}(\hat{\theta}) = 1/W_H$ has a (small) negative bias because the uncertainty about the study weights w_i is ignored. However, this bias is in practice negligible.

The effect of the study-level variance ω on the weights $w_i(\omega)$ is to reduce their dispersion and shrink their relative weights w_i/W_H toward the common value $1/H$. Therefore the effect of uncertainty about the missing value of a sampling variance diminishes with increasing ω. So, the case of $\omega = 0$, explored in the rest of the article represents an extreme case, albeit without taking the uncertainty about ω into account.

3.3. Examples

The fit of the model for the variances σ_i^2, $i = 1,\ldots,4$, yields the estimates $\hat{\alpha} = 141.23$ and $\hat{\gamma} = 18.60$, with estimated sampling variance matrix

$$\begin{pmatrix} 3870.33 & 611.86 \\ 611.86 & 103.93 \end{pmatrix}.$$

The estimated correlation of the two estimators is 0.965. The empirical Bayes estimate of the expected value of $\hat{\sigma}_5^2$ is $141.23/17.60 \times (1 + 2 \times 18.60/92) = 11.27$. The corresponding estimate of $\hat{\tau}_5^2$ is $11.27 \times 2/92 = 0.245$. By substituting this value for $\hat{\tau}_5^2$ we obtain the estimates $\hat{\theta}_+ = 0.382$, 0.499 and 0.445 in the respective cases A, B and C, each with estimated standard error 0.232. The latter is an underestimate in all three cases because we have pretended $\hat{\tau}_5^2$ to be known.

The uncertainty about $\hat{\tau}_5^2$ is partly reflected by averaging the plausible estimates $\tilde{\theta}_+$ obtained by substituting for $\hat{\tau}_5^2$ random draws from its fitted sampling (or posterior) distribution. The estimate of θ is obtained as the average of the plausible estimates. The sampling variance has two components: average of the plausible sampling variances and variance of

the plausible estimates of θ. The latter component should be multiplied by $(1+1/m)$; when we choose a large m, this factor makes next to no difference.

We applied $m = 1000$ replications; we can be profligate with the choice of m because the calculations that follow are simple. The averages in the three cases are 0.383, 0.500 and 0.445, and the standard errors are estimated by 0.232 in all three cases. Thus, the results are altered only slightly by using plausible values $\tilde{\tau}_5^2$. In fact, the estimates of the standard errors are greater by less than 0.0002 compared to when $\hat{\tau}_5^2$ is used.

The two kinds of imputation we applied are *improper* in the terminology of Rubin (2004) because they fail to reflect the uncertainty about the missing value(s) in its entirety. Specifically, we have pretended that the parameters α and γ were known and were equal to their estimates. We make amends for this by drawing a random sample of plausible pairs $(\tilde{\alpha}, \tilde{\gamma})$, and then drawing a plausible value $\tilde{\sigma}_{H+1}^2$ from each (plausible) distribution based on the realised pair $(\tilde{\alpha}, \tilde{\gamma})$. In this procedure, we assume that the sampling distribution of $(\hat{\alpha}, \hat{\gamma})$ is bivariate normal, with the sampling variance derived from the fitted information matrix. Relying on asymptotic normality with such a small sample size $H = 4$ is clearly problematic, and some error is committed. However, this is bound to be not as large as if we pretended this variance matrix to vanish.

With this method of multiple imputation, we obtain the estimates 0.378, 0.497 and 0.443 for the respective cases A, B and C, with standard error 0.234 in each case. They do not differ materially from the results obtained by simpler improper methods of imputation. The estimate of the standard error is inflated by only 0.0025.

In generating replicates of $(\tilde{\alpha}, \tilde{\gamma})$, we rejected 32 pairs because they contained at least one negative value. The values of the plausible weight $\tilde{w}_5$ ranged from 0.01 to 15.6; their mean was 4.07 and standard deviation 1.31. The substantial uncertainty about the parameters α and γ translates to substantial uncertainty about τ_5^2 or the weight w_5, but this does not contribute substantially to the uncertainty about θ.

4. Plausible values of $\hat{\theta}$

An approach that involves relatively weak assumptions about the incompletely reported study $H+1$ is based on a plausible range of values of τ_{H+1}^2. A plausible range, an interval $(\tau_{H+1,\mathrm{L}}^2, \tau_{H+1,\mathrm{U}}^2)$, is defined by the condition that all values of τ_{H+1}^2 outside this interval can be ruled out. An interval that subsumes a plausible range is also a plausible range, but a subinterval of a plausible range may not be. In ideal circumstances, the plausible range would be elicited from subject matter experts, such as clinical personnel involved in the meta-analysis or one of its studies. We assume that a plausible range for τ_{H+1}^2 has been specified. A value is said to be plausible if it is contained in the plausible range.

We evaluate the estimator $\hat{\theta}$ conditionally on τ_{H+1}^2 being equal to each value on a fine grid that covers the plausible range. These values can be regarded as plausible for $\hat{\theta}$, and the range they cover, $(\hat{\theta}_\mathrm{L}, \hat{\theta}_\mathrm{U})$, as the plausible range for $\hat{\theta}_+$. The plausible values of $\mathrm{var}(\hat{\theta}_+) = 1/(W_H + w_{H+1})$ can be established similarly. If these two plausible ranges are narrow, then we can conclude the analysis with these two intervals, admitting the uncertainty about $\hat{\theta}_+$ additional to its sampling variation, as well as the uncertainty about the standard error.

The results of the meta-analyses for the three sets of studies introduced in Table 1 are presented in the panels in one row of Figure 1. The first panel at the top (Ae, for case A) presents the plausible value of $\hat{\theta}_+$ as a function of the plausible value of $\hat{\tau}_5^2$ (solid line). The estimate $\hat{\theta}_-$ is indicated by horizontal dashes. By including study 5 in the meta-analysis, the estimate of $\hat{\theta}$ is increased by about 0.10, from $\hat{\theta} = 0.295$ to between $0.373 - 0.409$. The plausible standard errors, plotted in panel As are in the range $0.222 - 0.235$, reduced from the standard error of $\hat{\theta}_-$ equal to 0.263. The t ratio, plotted in panel At, is increased from 1.12 (based on $\hat{\theta}_-$) to between 1.59 and 1.85. So, we conclude with no evidence of a treatment effect, despite an appreciable increase in the estimate of θ and reduction in its standard error.

Panels in the middle row, based on case B, present an example in which study 5 alters the verdict of significance unequivocally, for any plausible value of $\hat{\tau}_5^2$. Panels at the bottom (case C) display an example of *impasse*. As a function of the estimate $\hat{\tau}_5^2$, the t ratio (panel Ct) intersects the horizontal line drawn at 1.96. There would be sufficient evidence of a treatment effect for some plausible values of $\hat{\tau}_5^2$, namely in the range (0.170, 0.217), but 'not significant' would be the verdict for $0.217 < \hat{\tau}_5^2 < 0.240$.

4.1. Plausible verdicts of hypothesis testing

If establishing significance is the sole objective of the analysis, then we can conclude the analysis with an unequivocal statement when the test of the relevant hypothesis yields the same verdict for every plausible value of τ_{H+1}^2. This approach can be reduced to evaluating the t ratio at the limits $\tau_{H+1,\mathrm{L}}^2$ and $\tau_{H+1,\mathrm{U}}^2$ and at most one other point. We assume that the t test is used throughout, and that its assumptions are satisfied.

Let $\tilde{\tau}^2$ be a plausible value of τ_{H+1}^2 and $\tilde{\theta}$ and $\tilde{w}_{H+1}$ the corresponding values of $\hat{\theta}_+$ and w_{H+1}. In Appendix A we show that, except when $\hat{\theta}_{H+1} = 0$, the t ratio, $\tilde{\theta}/\sqrt{W_H + \tilde{w}_{H+1}}$, is either a unimodal or a monotone function of $\tilde{\tau}^2$; its extension to the entire real axis has a single extreme, $w_{H+1}^* = W_H(\hat{\theta}_-/\hat{\theta}_{H+1} - 2)$, and is monotone in the two intervals separated by w_{H+1}^*. If w_{H+1}^* lies outside $(w_{H+1,\mathrm{L}}, w_{H+1,\mathrm{U}})$, then the t ratio is a monotone function of $\tilde{w}_{H+1} = 1/\tilde{\tau}^2$ in this range, and so it suffices to evaluate it at the limits $w_{H+1,\mathrm{L}}$ and $w_{H+1,\mathrm{U}}$; these values of t delimit the plausible range of the t ratio. When w_{H+1}^* is contained in $(w_{H+1,\mathrm{L}}, w_{H+1,\mathrm{U}})$, the plausible values of the t ratio have the same sign throughout, and so their range is delimited by the t ratio evaluated at w_{H+1}^* and at either $w_{H+1,\mathrm{L}}$ or $w_{H+1,\mathrm{U}}$. The plausible range of p values is obtained straightforwardly, as the p value is a decreasing function of $|t|$.

An interesting case arises when $\hat{\theta}_-$ and $\hat{\theta}_{H+1}$ have opposite signs and $W_H\,\hat{\theta}_-/\hat{\theta}_{H+1}$ is a plausible value of w_{H+1}. Zero is now a plausible value of the t ratio, so the ratio may be both positive and negative. But significance of the t ratio would be plausible only in some esoteric settings with extremely wide plausible ranges of τ_{H+1}^2.

Apart from adopting the t statistic for the hypothesis $\theta = 0$, the only assumption we have made is about the plausible range for w_{H+1}. For its specification we have to rely on expert opinion formed by information from other studies and the nature of the variation of the outcome variable in the relevant population. Eliciting such opinion is far from trivial. Experts may be ill-at-ease and reticent to cooperate, being concerned that the integrity and veracity

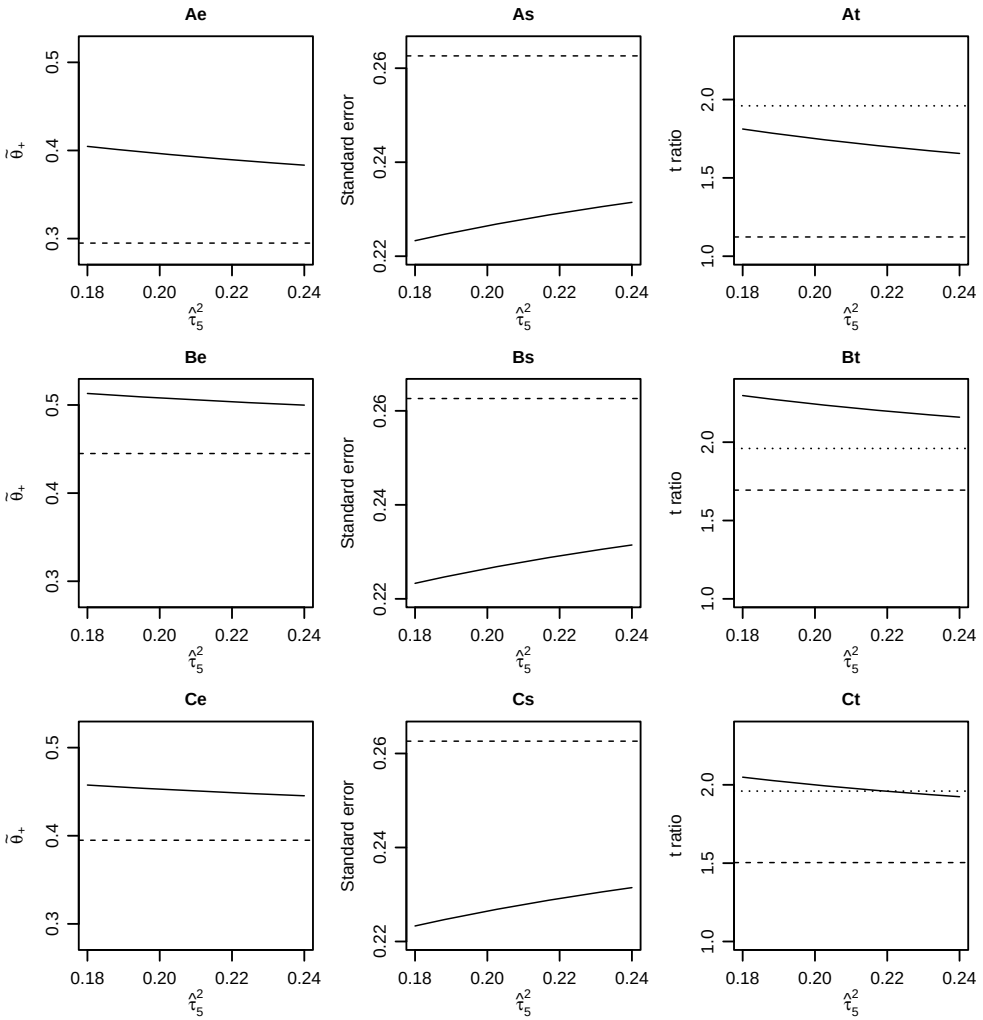

Figure 1: Plausible estimates (e), standard errors (s) and t ratios (t) in the meta-analyses A, B and C (Table 1). Horizontal dashes mark the relevant statistics for the four studies with complete information and the horizontal dots are drawn at the critical value of the t ratio, 1.96. By construction (identical sets of variance estimates $\hat{\tau}_i^2$ in Table 1), the panels As, Bs and Cs are identical.

of their statements may be undermined in the future when new information emerges.

An alternative to this approach involves finding the values of τ_{H+1}^2 for which the p value of 0.05, or another a priori selected value, is attained. The behaviour of the t ratio as a function of $\tau_{H+1}^2 = 1/w_{H+1}$ implies that there are at most two such values. These borderline points are easy to find by the Newton method or another line search algorithm. Within each interval delimited by a pair of these borderline values, the p value is either entirely greater or

smaller than the reference value, say 0.05. We then ask the experts whether every interval in which the p values are greater than 0.05 is entirely implausible. If the answer is affirmative, then we conclude with the verdict of significance because it would have been attained for *any* plausible value of τ^2_{H+1}.

A drawback of both approaches is the possibility of an impasse, which arises when significance would have been achieved for some but not all plausible values of τ^2_{H+1}. In such a case, we have to admit that both significance and its negation are plausible outcomes of the analysis. In general, it is preferable to specify as narrow a plausible range for τ^2_{H+1} as possible, to reduce the chances of an impasse. However, it is an imperative that any value of τ^2_{H+1} outside the declared range can be ruled out; otherwise the integrity of the method is breached. In the second variant of this approach, it is important to discourage a hasty or perfunctory dismissal of the plausibility of the intervals in which the p value is greater than the reference (0.05).

4.2. Accounting for the uncertainty about τ^2_{H+1}

The uncertainty that can be attributed to the unknown w_{H+1} is assessed by the conditional distribution of $\hat{\theta}_+$ given $\hat{\theta}_-$ and W_H. The Taylor expansion for $\hat{\theta}_+$ around $\hat{\theta}_-$ yields the approximation

$$\operatorname{var}\left(\hat{\theta}_+ \,|\, \hat{\theta}_-, \hat{\theta}_{H+1}\right) \doteq \left(\hat{\theta}_{H+1} - \hat{\theta}_-\right)^2 \mathrm{E}\left\{(1+r_{H+1})^{-4}\right\} \operatorname{var}(\hat{r}_{H+1}), \tag{2}$$

where $r_i = w_i/W_H$, $i = 1,\ldots,H+1$, is the relative weight. This identity is derived in Appendix B. It implies three factors that have an impact on the uncertainty about $\hat{\theta}_+$: the deviation of $\hat{\theta}_{H+1}$ from $\hat{\theta}_-$, the relative magnitude of w_{H+1} with respect to W_H, and the variance of this ratio r_{H+1}. The first factor does not involve r_{H+1}, and can be evaluated from the available data directly. It vanishes when $\hat{\theta}_{H+1} = \hat{\theta}_-$, and then $\hat{\theta}_+ = \hat{\theta}_-$ for any value of τ^2_{H+1}. The second term has an upper bound of 1.0. If study $H+1$ is large, then w_{H+1} is also large, and then this factor is small. Further, when we have a lot of information about θ, and so W_H is large, then $\operatorname{var}(\hat{r}_{H+1})$ is small. These considerations, however loose and involving approximation, conform with intuition. A study $H+1$ omitted from meta-analysis introduces greater uncertainty about θ when $\hat{\theta}_{H+1}$ is exceptional among the estimates $\hat{\theta}_1, \ldots, \hat{\theta}_H$, when the study contains a lot of information about θ (w_{H+1} is small), and when we are uncertain about w_{H+1}.

The first factor is equal to 0.167, 0.061 and 0.051 for the respective cases A – C presented in Table 1. The other two factors have values common to the three cases. The plausible values of r_{H+1} are in the range 0.265 – 0.406, and for $1/(1+r_{H+1})^4$ they are in the range 0.256 – 0.390. We approximate $\operatorname{var}(\hat{r}_{H+1})$ conservatively by the variance of the uniform distribution on $(0.265, 0.406)$, that is, $0.141^2/12 = 0.00166$, and the expectation of $1/(1+r_{H+1})^4$ by its largest plausible value, $1/1.265^4 = 0.390$. Thus a conservative estimate of the variance in (2) is $0.167 \times 0.00166 \times 0.391 = 1.08 \cdot 10^{-4}$, that is, standard deviation of about 0.0104 in case A, 0.0063 in case B, and 0.0057 in case C. This is a small contribution to the overall uncertainty attributable to the variation of the outcome variable in the studied population(s), as quantified by $1/\sqrt{W_H} = 0.263$.

Table 2: Estimates, standard errors and sample sizes ($n^{(\mathrm{P})}$ — placebo and $n^{(\mathrm{M})}$ — mirtazapine) for the studies in the systematic review of Mavridis *et al.* (2014).

Study	Estimate	St. error	$n^{(\mathrm{P})}$	$n^{(\mathrm{M})}$
S1	–3.1	2.91	18	25
S2	–2.5	2.20	23	26
S3	–1.8	3.02	16	11
S4	–6.8	2.30	23	22
S5	3.6	1.71	20	21
S6	–4.6	2.26	26	29
S7	–2.3	1.97	32	34
S8	–2.9	1.68	47	50

5. Example II

In this section we illustrate the methods on a systematic review conducted by Mavridis *et al.* (2014) for comparing mirtazapine, a drug for treating clinical depression, with placebo. The outcome variable is recorded on the HAMD21 scale constructed originally by Hamilton (1967) using a patient questionnaire. Larger values of HAMD21 correspond to more severe illness.

The systematic review found eight studies. Their results are presented in Table 2, listing the estimate of the treatment effect ($\hat{\theta}_i$), its (estimated) standard error, and the number of observations ($n_i^{(\mathrm{P})}$ for placebo and $n_i^{(\mathrm{M})}$ for mirtazapine) for each study S1 – S8. Study S5, the only one with $\hat{\theta}_i > 0$, is an obvious outlier. The standard errors are in the range 1.68–3.02, and the numbers of observations are in the range 27–97.

We pretend that one of the standard errors is missing and apply the methods that make use of the estimate and sample size for this study. The results are presented in Table 3. Row labelled –Si, $i = 1, \ldots, 8$, represents the setting with the standard error in study Si missing. The first two columns present the estimates of the parameters of the inverse gamma distribution on which imputation for this missing value is based. The next column presents the estimate of the treatment effect based on the seven retained studies ($\hat{\theta}_-$). The next two columns present the imputed standard error $\tilde{\tau}_i$ derived from the (empirical) posterior expectation of the variance σ_i^2 and the estimate of θ based on this standard error ($\hat{\theta}_+$). The right-most column displays the estimates and standard errors based on averaging over 100 random draws from the posterior distribution of σ_i^2.

The results are presented with three decimal places, so that the small differences of the estimates can be discerned. The target of estimation is the treatment effect based on the estimates of all the eight studies, $\hat{\theta} = -2.061$. With no data discarded, the standard error of $\hat{\theta}$ is estimated by 0.751. By imputing the posterior mean, all single-imputation estimates $\hat{\theta}_+$ are close to the target, except for the setting –S5. The estimated standard errors are also close to 0.751, except for –S5. All of them should exceed 0.751 because they are based on less information. The contradiction arises because the uncertainty about the imputed

Table 3: Estimates and estimated standard errors (se) of the treatment effect; metaanalysis of studies summarised in Table 2, with one standard error $\hat{\tau}_i$ deleted; θ_- — the study discarded altogether; θ_+ — the study included, with the posterior expectation of σ_i imputed; $\tilde{\theta}_+$ — the study included, with multiple imputation for σ_i.

Row labelled –Sk, $k = 1, \ldots, 8$, indicates that $\hat{\tau}_k$ is regarded as missing.

	$\hat{\alpha}$	$\hat{\gamma}$	$\hat{\theta}_-$ (se)	$\tilde{\tau}_i$	$\hat{\theta}_+$ (se)	$\tilde{\theta}_+$ (se)
–S1	3.03	7.46	–1.987 (0.777)	2.36	–2.096 (0.748)	–2.098 (0.753)
–S2	2.89	6.63	–2.003 (0.799)	2.23	–2.060 (0.752)	–2.061 (0.758)
–S3	3.05	6.87	–2.079 (0.775)	3.26	–2.064 (0.754)	–2.063 (0.764)
–S4	2.89	6.62	–1.496 (0.795)	2.34	–2.043 (0.752)	–2.050 (0.767)
–S5	5.54	13.45	–3.414 (0.836)	2.59	–2.750 (0.795)	–2.753 (0.797)
–S6	2.94	7.00	–1.746 (0.796)	2.05	–2.120 (0.752)	–2.121 (0.753)
–S7	2.93	6.92	–2.021 (0.812)	1.85	–2.066 (0.745)	–2.064 (0.754)
–S8	3.09	7.57	–1.852 (0.840)	1.47	–2.108 (0.739)	–2.113 (0.748)

standard error is ignored. Multiple imputation corrects this bias but the estimates $\tilde{\theta}_+$ differ from their single-imputation counterparts $\hat{\theta}_+$ only slightly.

The estimates stand out for the setting –S5 because study S5 has a smaller standard error than its sample size suggests. When $\tilde{\tau}_5$ is imputed the contribution of S5 to estimating θ is underrated, and so its influence is reduced. In summary, imputation of the posterior mean of the variance is sufficient for estimating the treatment effect. Multiple imputation yields similar estimates and inflates the standard errors only slightly.

We illustrate sensitivity analysis by pretending that $\hat{\tau}_5$ is not recorded. Since $\hat{\theta}_5$ is an outlier among the estimates, it may be justified to discard the study altogether, especially if a careful review of the literature and of other sources discovers some reason for the exceptional result. Failure to report $\hat{\tau}_5$ might also raise suspicion about both the quality and context of the study. Instead of the dichotomy, to include or exclude the study from the meta-analysis, we define a plausible range of standard errors, $(\tilde{\tau}_{5\mathrm{L}}, \tilde{\tau}_{5\mathrm{U}})$. Exclusion corresponds to $\tilde{\tau}_{5\mathrm{L}} = +\infty$, implying that also $\tilde{\tau}_{5\mathrm{U}} = +\infty$. Exclusion being plausible corresponds to $\tilde{\tau}_{5\mathrm{L}} < +\infty$ and $\tilde{\tau}_{5\mathrm{U}} = +\infty$.

An analyst might impute for $\hat{\tau}_5$ the standard error from a study with a similar sample size, such as S1 or S4, and allowing some larger values to reflect the doubt about $\hat{\theta}_5$. Suppose the plausible range for $\hat{\tau}_5$ is set to $(2,5)$. Figure 2 displays the plot of the plausible values of $\hat{\theta}_+$ as a function of $\tilde{\tau}_5$ (solid line) together with the plausible confidence intervals (shaded area). For $\tilde{\tau}_5 = 2.0$, $\hat{\theta}_+$ is close to the target $\hat{\theta}_+ = -2.06$ (horizontal line of long dashes), so the error caused by the failure to allow for $\hat{\tau}_5 = 1.7 < \tilde{\tau}_{5\mathrm{L}}$ is not harsh.

For $\tau_5 = 5.0$, study S5 contributes with very small weight; $\hat{\theta}_+$ is close to $\hat{\theta}_-$, marked by the horizontal dashed line. The standard error increases with $\tilde{\tau}_5$, from 0.771 at $\tilde{\tau}_5 = 2.0$ to 0.824 at $\tilde{\tau}_5 = 5$ — the grey region narrows towards the right. The upper confidence limit for $\hat{\theta}_+$ decreases with $\tilde{\tau}_5$. It crosses zero at $\tilde{\tau}_5 = 1.32$. So, there is evidence of a negative effect of mirtazapine so long as $\hat{\tau}_5 > 1.32$. If $\hat{\tau}_5$ were very small, study S5 would dominate

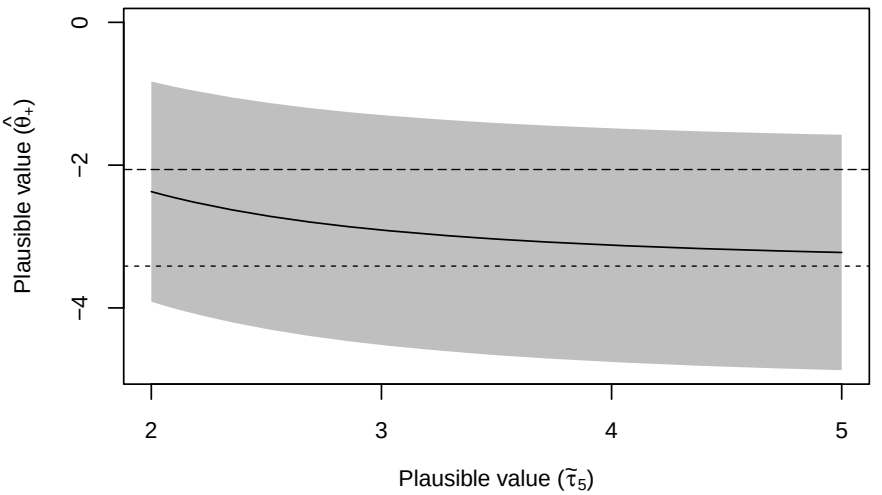

Figure 2: Sensitivity analysis. Plausible estimate of the treatment effect $\hat{\theta}$ as a function of the plausible value of the standard error τ_5 when $\hat{\tau}_5$ for study S5 is masked, regarded as not reported.

the meta-analysis and it would conclude with evidence of a positive (detrimental) effect of mirtazapine. This would happen for $\hat{\tau}_5 < 0.64$. Such an outcome would not be credible given that all but one study yielded a negative estimate.

6. Discussion

The empirical Bayes approach in Section 3 involves some assumptions that are contentious and their plausibility is difficult to assess. The approach motivated by sensitivity analysis in Section 4 carries a lighter burden of assumptions but has a heavier demand on input — it requires the declaration of a plausible range for the missing sampling variance. Also, it may conclude with an impasse, when one conclusion, e.g., of a hypothesis test, is obtained for some plausible values of this variance, and another conclusion for other values that are equally plausible.

The analysis in Section 2 shows that a study with standard error not reported can contribute to the estimation of the overall treatment effect. Accounting for the uncertainty about the imputed standard error makes much less difference. Similar conclusions can be drawn about using plausible values for the missing standard error(s).

The model applied in Section 3 can be expanded to a regression model, and thus strengthen the inference about a missing variance by exploiting the association of the variance and mean implied by the distribution of the outcomes or other auxiliary information. This approach is not always useful. For example, when the outcomes are binary and the events

are neither rare nor very frequent the variance is a very flat function of the probability. In any case, the small number of studies precludes any complex modelling and any reliable inference about the mean-variance relationship; prior information may be more useful.

The examples in Sections 4 and 5 confirm that even simple methods, using some shortcuts on proper imputation, exploit nearly to the full the information about an incompletely reported study and they estimate the standard error of the overall treatment effect with negligible bias. Sensitivity analysis using plausible ranges for the missing variance has some potential but this is undermined by the general reluctance to participate in elicitation of these ranges.

The common or average treatment effect θ is ascribed importance, and motivates the attempt to recover information contained in an incompletely reported study $(H+1)$. The problem has some commonality with publication bias, a widely studied issue. The expectation θ can be interpreted as the treatment effect in a set of studies among which the realised studies are a random sample. This interpretation has a flaw in that the treatment effects of these studies, $\theta_1, \ldots, \theta_{H+1}$, would be a random sample from a meaningful distribution only if the populations (constituencies) and contexts of the studies were selected at random from a universe of potential studies, that is, according to a design. In practice, these aspects are selected haphazardly, influenced by the availability of expertise and funding and concern about the specific issue. Also, the contexts of the realised studies, especially those conducted in the more distant past and in countries with different levels of development and organisation of health care, may differ a great deal from the context for which the inferences drawn by a meta-analysis are intended. Such relevance is rarely incorporated in the weights used for estimating the overall (or average) treatment effect.

All the data used in this article are displayed in Tables 1 and 2.

References

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schulz, K.F., Simel, D., and Stroup, D.F., (1996). Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *Journal of the American Medical Association*, 276, pp. 637–639.

DerSimonian, R., and Laird, N., (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7, pp. 177–188.

Duval, S.J., and Tweedie, R.J., (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, pp. 455–463.

Gamble, C., and Hollis, S., (2005). Uncertainty method improved on best-worst case analysis in a binary meta-analysis. *Journal of Clinical Epidemiology*, 58, pp. 579–588.

Glass G.V., (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5, pp. 3–8.

Haidich, A.B., (2010). Meta-analysis in medical research. *Hippokratia*, 14, pp. 29–37.

Hamilton, M., (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology*, 6, pp. 278–296.

Hedges L.V., and Olkin I., (1985). *Statistical Methods for Meta-Analysis.* Boston, MA: Academic Press.

Lin, L., and Chu, H., (2018). Quantifying publication bias in metaanalysis. *Biometrics*, 74, pp. 785–794.

Longford, N.T., (2010). Small-sample inference about variance and its transformations. *SORT, Journal of the Catalan Institute of Statistics*, 34, pp. 3–29.

Longford, N.T., (2015). On the inefficiency of the restricted maximum likelihood. *Statistica Neerlandica*, 69, pp. 171–196.

Mavridis, D., White, I.R., Higgins, J.P.T., Cipriani, A., and Salanti, G., (2014). Allowing for uncertainty due to missing continuous outcome data in pairwise and network meta-analysis. *Statistics in Medicine*, 34, pp. 721–741.

Rice, K., Higgins, J.P.T., and Lumley, T., (2018). A re-evaluation of fixed effect(s) meta-analysis. *Journal of the Royal Statistical Society* Series A, 181, pp. 205–227.

Rothstein, H.R., Sutton, A.J., and Borenstein, M., (2005). *Publication Bias in Meta-Analysis. Prevention, Assessment and Adjustments.* Wiley and Sons, Chichester, UK.

Rubin, D.B., (2004). *Multiple Imputation for Nonresponse in Surveys.* 2nd ed. New York, NY: Wiley.

von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C., and Vanderbroucke, J.P., (2008). STROBE Initiative. The strengthening of reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Journal of Clinical Epidemiology*, 61, pp. 344–349.

Appendix A. The t ratio for $\hat{\theta}$ as a function of w_{H+1}

In this appendix we explore the behaviour of the t ratio $\hat{\theta}_+/\sqrt{W_H + w_{H+1}}$ as a function of w_{H+1}. This function is

$$T(w) = \frac{W_H\,\hat{\theta}_- + w\hat{\theta}_{H+1}}{\sqrt{W_H + w}}\,.$$

Its derivative is

$$\begin{aligned}\frac{\partial T}{\partial w} &= \frac{1}{\sqrt{W_H + w}}\left(\hat{\theta}_{H+1} - \frac{1}{2}\,\frac{W_H\,\hat{\theta}_- + w\hat{\theta}_{H+1}}{W_H + w}\right)\\ &= \frac{1}{2\,(W_H + w)^{\frac{3}{2}}}\left\{W_H\left(2\hat{\theta}_{H+1} - \hat{\theta}_-\right) + w\hat{\theta}_{H+1}\right\}.\end{aligned}$$

The sign of this derivative does not depend on w when $\hat{\theta}_{H+1} = 0$. In that case, the derivative has the same sign as $-\hat{\theta}_-$. When $\hat{\theta}_{H+1} \neq 0$, the derivative has a single root at

$$w^*_{H+1} = W_H\left(\frac{\hat{\theta}_-}{\hat{\theta}_{H+1}} - 2\right),$$

where its sign switches from positive to negative or vice versa. Therefore T changes at w^*_{H+1} from decreasing to increasing or vice versa. Its value at w^*_{H+1} is

$$T\left(w^*_{H+1}\right) = \frac{2\,W_H}{\sqrt{W_H + w^*_{H+1}}}\left(\hat{\theta}_- - \hat{\theta}_{H+1}\right).$$

When $\hat{\theta}_{H+1} \neq 0$, the function T has a single root at $w^{(0)}_{H+1} = -W_H\,\hat{\theta}_-/\hat{\theta}_{H+1}$, which is positive when $\hat{\theta}_-$ and $\hat{\theta}_{H+1}$ have opposite signs. In that case, $w^*_{H+1} < 0$, and so T is monotone in the plausible range. In summary, T is either unimodal without changing its sign in the plausible range of w_{H+1}, or is monotone, in which case it may cross zero at one point.

Appendix B. Taylor expansion for $\hat{\theta}_+$

The first-order partial differential of $\hat{\theta}_+$ with respect to w_{H+1} is

$$\begin{aligned}\frac{\partial\hat{\theta}_+}{\partial w_{H+1}} &= \frac{(W_H + w_{H+1})\,\hat{\theta}_{H+1} - W_H\,\hat{\theta}_- - w_{H+1}\,\hat{\theta}_{H+1}}{(W_H + w_{H+1})^2}\\ &= \frac{W_H\left(\hat{\theta}_{H+1} - \hat{\theta}_-\right)}{(W_H + w_{H+1})^2}\\ &= \frac{\hat{\theta}_{H+1} - \hat{\theta}_-}{W_H}\,\frac{1}{\left(1 + \frac{w_{H+1}}{W_H}\right)^2},\end{aligned}$$

from which the expression for the conditional variance in equation (2) follows directly, evaluating $(\partial\hat{\theta}_+/\partial w_{H+1})^2\,\mathrm{var}(\hat{w}_{H+1})$ and substituting $r_{H+1} = w_{H+1}/W_H$.