Journal of the
Royal Statistical Society
SERIES B
Statistical
Methodology

J. R. Statist. Soc. B (2013)
75, *Part* 3, *pp.* 451–469

Estimating heterogeneity variance in meta-analysis

Andrew L. Rukhin

National Institute of Standards and Technology, Gaithersburg, USA

[Received August 2011. Final revision August 2012]

Summary. Several new estimators of the between-study variability in a heterogeneous random-effects meta-analysis model are derived. One is the unbiased statistic which is locally optimal for small values of the parameter. Others are Bayes procedures within a class of quadratic statistics for a diffuse prior with different choices of the prior mean. These estimators are compared with the DerSimonian–Laird procedure and the Hedges statistic in particular via the quadratic risk of the treatment effect estimator. A Monte Carlo study supports the usage of confidence intervals derived from the new estimators.

Keywords: Bayes estimators; Diffuse prior; Heteroscedasticity; Interlaboratory study; Random-effects model; Unbiased estimators

1. Introduction and summary

In many applications of meta-analysis it is important to assess the degree of heterogeneity among several participating studies, methods, instruments or laboratories which are supposed to measure the difference between two treatments or properties of the same material. Commonly this heterogeneity is non-negligible and cannot be easily explained in terms of available covariates. When one has to combine data under a random-effects heterogeneous model, the estimation of the between-study variance is essential.

Assume that there are n_i observations in the study i, $i=1,\ldots,p$. The data x_{ik}, $k=1,\ldots,n_i$, in this study are supposed to have the form

$$x_{ik}=\mu+l_i+\varepsilon_{ik}, \tag{1}$$

where μ is the treatment effect or the property value and l_i represents the study (or method) effect with zero mean and unknown heterogeneity variance σ^2. The independent zero-mean random errors ε_{ik} have possibly different variances τ_i^2. For a fixed i, the mean of $x_i=\Sigma_k x_{ik}/n_i$ is μ, and its variance is $\sigma^2+\sigma_i^2$, where $\sigma_i^2=\tau_i^2/n_i$. It is commonly assumed that ls and εs have Gaussian distributions. Then the classical statistic

$$s_i^2=\frac{\sum_k (x_{ik}-x_i)^2}{n_i(n_i-1)}$$

with the distribution $\sigma_i^2\chi^2(\nu_i)/\nu_i$, $\nu_i=n_i-1$, estimates σ_i^2 unbiasedly and is independent of x_i and s_j^2, $j\neq i$.

Commonly, the full data set is not available and only summary statistics x_i and s_i^2 are provided. In interlaboratory testing applications which stimulated this work, p is not large, whereas n_i may not be even given.

Address for correspondence: Andrew L. Rukhin, Statistical Engineering Division, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, USA.
E-mail: andrew.rukhin@nist.gov

If $\sigma^2+\sigma_i^2$ were known (up to a factor), then the best linear unbiased estimator of μ, $\tilde{\mu}=\Sigma_i\,\omega_i x_i$ with normalized weights

$$\omega_i=\frac{(\sigma^2+\sigma_i^2)^{-1}}{\sum_j(\sigma^2+\sigma_j^2)^{-1}},$$

could be used. Since the variances are unknown, these optimal weights must be estimated, and this problem leads to estimation of σ^2 which gains even more importance from the necessity of estimating the variance of $\tilde{\mu}$ and from confidence intervals for μ. In fact the parameter σ^2 has a meaning of its own. According to Higgins *et al.* (2009), page 139,

> 'The naive presentation of inference only on the mean of the random-effects distribution is highly misleading. Estimation of σ^2 is just as important.'

In some situations σ^2 can be treated as a possibly negative variance component such that $\sigma^2+\sigma_i^2\geqslant 0$. However, in the applications mentioned σ^2 always has the meaning of variance so that it is non-negative. Its estimation presents substantial difficulties, there are no positive unbiased estimators of this parameter, the Fisher information about it in model (1) is typically small, the explicit form of the Bayes procedures or of the (restricted) maximum likelihood estimator is lacking, its numerical evaluation may be surprisingly delicate (Böhning *et al.*, 2004), etc.

For this reason we introduce in Section 2 a quite natural class of simple quadratic estimators of σ^2. The formulae for the mean-squared error are derived, and locally optimal estimators which minimize the quadratic risk within this class at a given parameter value are determined in Sections 3 and 4. One of the main findings of this work is in the central for this paper Section 5. It is the approximate Bayes estimator of the between-studies variance σ^2 within the class considered. This estimator has a very explicit form which is completely determined by specification of the prior mean. Section 7 provides examples of all these procedures when $p=2$, and Section 10 gives an example of a heterogeneous air flow study.

The only features of the normal distribution that is used in this paper to derive σ^2-estimators are the formula for the kurtosis of a normal variable, the χ^2-distribution of s_i^2 and its independence from x_i. Each of the last two conditions is known to characterize the normal distribution but, if $\sigma_i^2=s_i^2$ are treated as given constants, the normality assumption is not needed. This situation is discussed in Section 6. Confidence intervals for μ derived by using the estimators obtained for σ^2 are discussed in Section 8. These intervals are compared via a Monte Carlo study in Section 9. Appendix A contains all mathematical derivations.

2. Variance estimators

We consider here the class of σ^2-estimators which are linear functions of quadratic statistics. This class includes the popular meta-analytic procedures of DerSimonian and Laird (1986) and of Hedges (1983).

The first of these methods estimates the heterogeneity variance as

$$\tilde{\sigma}^2_{\mathrm{DL}}=\frac{\sum_i(x_i-\tilde{x}_{\mathrm{GD}})^2 s_i^{-2}-p+1}{\sum_i s_i^{-2}-\sum_i s_i^{-4}/\sum_i s_i^{-2}},$$

where $\tilde{x}_{\mathrm{GD}}=\Sigma_i\,x_i s_i^{-2}/\Sigma_i\,s_i^{-2}$ is one of the traditional estimators of the common mean (the Graybill–Deal estimator).

The resulting DerSimonian and Laird (1986) estimator for μ,

$$\tilde{x}_{\mathrm{DL}} = \sum_i \frac{x_i}{\max(0, \tilde{\sigma}^2_{\mathrm{DL}}) + s_i^2} \Big/ \sum_i \frac{1}{\max(0, \tilde{\sigma}^2_{\mathrm{DL}}) + s_i^2},$$

is very popular in biostatistics, presenting an efficient alternative to the (restricted) maximum likelihood procedure. It admits extensions to more general models than model (1) (Böhning *et al.*, 2002). Several approximations to the Wald statistic for $\tilde{x}_{\mathrm{DL}}$ as a t-distribution with degrees of freedom ranging from $p-4$ to $p-1$ have been suggested (Higgins *et al.* (2009), section 5.2).

Another interesting rule,

$$\tilde{\sigma}^2_{\mathrm{H}} = \frac{\sum_i (x_i - \bar{x})^2}{p-1} - \frac{\sum_i s_i^2}{p},$$

is an analysis-of-variance motivated estimator, which was first employed in meta-analysis by Hedges (1983) and rediscovered in DerSimonian and Kacker (2007).

We unify these two important meta-analysis procedures in one class which requires knowledge only of x_i and s_i^2. Namely, the estimators under consideration have the form

$$\tilde{\sigma}^2 = \sum_{1 \leqslant i < j \leqslant p} c_{ij}(x_i - x_j)^2 + \sum_i d_i s_i^2 \tag{2}$$

with some constants c_{ij} and d_i. Thus class (2) consists of all linear functions of the $p(p+1)/2$-dimensional random vector Y with co-ordinates $(x_i - x_j)^2$, $i < j$, and s_i^2. Under the normality assumption, Y forms a sufficient (incomplete for $p \geqslant 3$) statistic for the parameters of interest.

According to formula (24) in Appendix A.1,

$$E(\tilde{\sigma}^2) = c\sigma^2 + \sum (c_i + d_i)\sigma_i^2, \tag{3}$$

with $c = \Sigma_{ij} c_{ij}$ and $c_i = \Sigma_j c_{ij}$, under the convention $c_{ij} = c_{ji}$ and $c_{ii} = 0$. Thus if $c = 1$ and $c_i = -d_i$, $\tilde{\sigma}^2$ is unbiased.

The method-of-moments estimator arises from equating the observed value of $\tilde{\sigma}^2$ to the right-hand side of equation (3) where σ_i^2 are replaced by s_i^2, i.e.

$$\tilde{\sigma}^2_{\mathrm{M}} = \frac{\sum_{1 \leqslant i < j \leqslant p} c_{ij}(x_i - x_j)^2 - \sum_i c_i s_i^2}{c}.$$

Thus the method-of-moments estimator $\tilde{\sigma}^2_{\mathrm{M}}$ is unbiased and has the form (2) with $\Sigma c_{ij} = 1$ and $d_i = -c_i$.

Both DerSimonian–Laird and Hedges variance estimators are derived from the method of moments, i.e. they are representable as $\tilde{\sigma}^2_{\mathrm{M}}$ for some coefficients c and d. To see that let, for positive weights $w_1, \ldots, w_p$,

$$T_m = \sum_i w_i^m, \qquad m = 1, \ldots,$$

and denote the Kronecker symbol by δ_{ij}: $\delta_{ij} = 1$, if $i = j$; $\delta_{ij} = 0$ otherwise. Then the choice $w_i = s_i^{-2}$ and $c_{ij} = (1 - \delta_{ij}) w_i w_j / (T_1^2 - T_2)$ produces the DerSimonian–Laird estimator. Indeed according to Lagrange's identity (Beckenbach and Bellman (1961), chapter 1, section 3), for any $w_1, \ldots, w_p$,

$$\sum_{1 \leqslant i < j \leqslant p} w_i w_j (x_i - x_j)^2 = \sum_i w_i \sum_i w_i \left(x_i - \frac{\sum_j w_j x_j}{\sum_j w_j} \right),$$

so, for $\tilde{\sigma}^2_{\mathrm{DL}}$, $d_i = -c_i = w_i(w_i - T_1)/(T_1^2 - T_2)$, $c = 1$, confirming its unbiasedness. The Hedges estimator corresponds to $c_{ij} \equiv \{(p-1)p\}^{-1}$ and $w_i = c_i = -d_i = p^{-1}$ in equation (2).

If, as commonly happens in meta-analysis, the σ_i^2 are supposed to be known (in which case they are replaced by given s_i^2), class (2) consists of statistics $\Sigma_{1 \leqslant i<j \leqslant p} c_{ij}(x_i - x_j)^2 + d$ for some c_{ij} and d. Dangers of treating s_i^2 as known, whereas they are in fact random, are discussed by Böhning and Sarol (2000).

The unbiased estimators of σ^2 cannot be positive. Indeed, if δ were such an estimator, then its expected value for $\sigma^2 = 0$ must vanish, implying that, with probability 1, $\delta \equiv 0$, which is not an unbiased statistic. If some of the cs or ds are negative, a non-negative variance estimator is obtained from equation (2) by truncating it at zero, $\tilde{\sigma}_+^2 = \max(0, \tilde{\sigma}^2)$. This estimator has a positive bias. However, the corresponding estimator of μ, the weighted mean,

$$\tilde{x} = \sum_i x_i (\tilde{\sigma}_+^2 + s_i^2)^{-1} \Big/ \sum_i (\tilde{\sigma}_+^2 + s_i^2)^{-1}$$

is unbiased provided that s_i^2 are independent of x_i.

As already noted, when $c_{ij} = (1 - \delta_{ij}) w_i w_j$ for fixed positive weights $w_1, \ldots, w_p$, we have

$$\sum_{1 \leqslant i<j \leqslant p} c_{ij}(x_i - x_j)^2 = T_1 \sum_i w_i (x_i - \tilde{\mu})^2$$

with $\tilde{\mu} = \Sigma_i w_i x_i / T_1$. If $c_{ij} = (1 - \delta_{ij}) w_i w_j (w_i + w_j)$ then, for the same $\tilde{\mu}$,

$$\sum_{1 \leqslant i<j \leqslant p} c_{ij}(x_i - x_j)^2 = T_1 \sum_i w_i^2 (x_i - \tilde{\mu})^2 + T_2 \sum_i w_i (x_i - \tilde{\mu})^2.$$

Thus the natural class of fairly simple estimators (2), besides the classical by now procedures, includes for any weights w not only the traditional Cochran heterogeneity statistic $\Sigma_i w_i (x_i - \tilde{\mu})^2$, but also the statistic $\Sigma_i w_i^2 (x_i - \tilde{\mu})^2$ appearing in the maximum likelihood analysis (Rukhin and Vangel, 1998) and in the form of the best unbiased estimator (Section 3).

One of the estimators not belonging to class (2) (being a non-linear function of Y) is the empirical Bayes estimator that was proposed by Morris (1983). For $p \geqslant 3$, this procedure $\tilde{\sigma}^2_{\mathrm{MP}}$ can be defined as a solution to the equation

$$\sum_{1 \leqslant i<j \leqslant p} \frac{(x_i - x_j)^2}{(\sigma^2 + s_i^2)(\sigma^2 + s_j^2)} = (p-1) \sum_i \frac{1}{\sigma^2 + s_i^2},$$

which can be shown to be equivalent to the method that was introduced by Mandel and Paule (1970) or an approximation to the restricted maximum likelihood rule.

3. Best unbiased estimator

Biggerstaff and Tweedie (1997) have shown that the variance of $\tilde{\sigma}^2_{\mathrm{DL}}$ has the form

$$\kappa_0^2 = \frac{2\sigma^4 (T_1^2 T_2 - 2T_1 T_3 + T_2^2) + 4\sigma^2 (T_1^3 - T_1 T_2) + 2(p-1) T_1^2}{(T_1^2 - T_2)^2}. \tag{4}$$

A different interpretation of this formula is $E\{\mathrm{var}(\tilde{\sigma}^2_{\mathrm{DL}} | s_1^2, \ldots, s_p^2)\} = \kappa_0^2$.

The variance of the Hedges estimator for fixed $\sigma_i^2 = s_i^2$ is

$$\kappa_\infty^2 = \frac{2\sigma^4 p(p-1) + 4\sigma^2 (p-1) \sum \sigma_i^2 + 2(p-2) \sum \sigma_i^4 + 2p^{-1} \left(\sum \sigma_i^2\right)^2}{p(p-1)^2}, \tag{5}$$

e.g. Friedman (2000). Note that, for any weights w,

$$\frac{1}{p-1} \leqslant \frac{T_1^2 T_2 - 2T_1 T_3 + T_2^2}{(T_1^2 - T_2)^2} \leqslant 1, \tag{6}$$

which implies that for large σ^2 the variance of the Hedges estimator is smaller than that of the DerSimonian–Laird estimator (strictly so if $p>2$). This fact is well known from simulation results (e.g. Viechtbauer (2005)).

Our result in Appendix A.1 unifies equations (4) and (5). Namely, it is shown that, for any estimator $\tilde{\sigma}^2$ of the form (2),

$$\begin{aligned} \tfrac{1}{2} E\{\mathrm{var}(\tilde{\sigma}^2 | s_1^2, \ldots, s_p^2)\} = \sigma^4 \left(\sum_i c_i^2 + \sum_{i \neq j} c_{ij}^2 \right) + 2\sigma^2 \left(\sum_i c_i^2 \sigma_i^2 + \sum_{i \neq j} c_{ij}^2 \sigma_j^2 \right) + \sum_i c_i^2 \sigma_i^4 \\ + \sum_{i \neq j} c_{ij}^2 \sigma_i^2 \sigma_j^2. \end{aligned} \tag{7}$$

Both formulae (4) and (5) follow from equation (7). Indeed in the notation of Section 2, if $c_{ij} = (1-\delta_{ij}) w_i w_j / (T_1^2 - T_2)$, then $c_i = w_i (T_1 - w_i)/(T_1^2 - T_2)$, and $c = \Sigma c_i = 1$.

Minimization in w_i of the coefficient at σ^4 gives $w_i = \{p(p-1)\}^{-1}$, i.e. leads to the Hedges estimator. However optimization of the σ^2-free term in equation (7) does not result in $w_i \propto \sigma_i^{-2}$, which would correspond to the DerSimonian–Laird procedure. This fact may explain the large mean-squared error of $\tilde{\sigma}^2_{\mathrm{DL}}$ which was observed in many simulations (Malzahn *et al.*, 2000; Jackson *et al.*, 2010) and which is discussed further.

Assuming that $\sigma^2 + \sigma_i^2$, $i = 1, \ldots, p$, are fixed, we find the form of the optimal unbiased estimator in class (2). Let K_1 be the covariance matrix of the $p(p-1)/2$-dimensional random vector with co-ordinates $(x_i - x_j)^2$, $i<j$. Because of the normality assumption, its diagonal elements are $2(2\sigma^2 + \sigma_i^2 + \sigma_j^2)^2$, and its non-zero off-diagonal elements, $2(\sigma^2 + \sigma_i^2)^2$, correspond to pairs which have one common index i. Then with $\mathbf{c}$ formed by co-ordinates c_{ij}, $i<j$, equation (7) can be written in the form

$$E\{\mathrm{var}(\tilde{\sigma}^2 | s_1^2, \ldots, s_p^2)\} = \mathbf{c}^{\mathrm{T}} K_1 \mathbf{c},$$

for unbiased estimators of σ^2, i.e. when $2e^{\mathrm{T}}\mathbf{c} = 1$. Here and elsewhere e is the $p(p-1)/2$-dimensional vector with unit co-ordinates.

The vector $(e^{\mathrm{T}} K_1^{-1} e)^{-1} K_1^{-1} e/2$ minimizes this quadratic form. For $w_i = (\sigma^2 + \sigma_i^2)^{-1}$, let $\mathbf{c}^0$ have co-ordinates, for $1 \leqslant i < j \leqslant p$,

$$c_{ij}^0 = \frac{w_i w_j \{(w_i + w_j) T_1 - T_2\}}{T_1^2 T_2 - 2T_1 T_3 + T_2^2}. \tag{8}$$

Appendix A.2 verifies that $K_1 \mathbf{c}^0 = 2T_1^2 (T_1^2 T_2 - 2T_1 T_3 + T_2^2)^{-1} e$, and

$$c_i^0 = \frac{w_i^2 (T_1^2 - 2w_i T_1 + T_2)}{T_1^2 T_2 - 2T_1 T_3 + T_2^2}.$$

Therefore, $2e^{\mathrm{T}}\mathbf{c}^0 = \Sigma_{i \neq j} c_{ij}^0 = 1$ and, for a fixed σ^2, equation (8) provides the coefficients of the best unbiased estimator in class (2),

$$\tilde{\sigma}^2_{\mathrm{U}} = \frac{T_1^2 \sum w_i^2 (x_i - \tilde{x})^2 - (T_1^2 + T_2) \sum w_i^2 s_i^2 + 2T_1 \sum w_i^3 s_i^2}{T_1^2 T_2 - 2T_1 T_3 + T_2^2}. \tag{9}$$

For equal weights $w_i \equiv 1$ and $\sigma^2 = \infty$, $\tilde{\sigma}^2_{\mathrm{U}}$ coincides with the Hedges estimator. If $\sigma^2 = 0$, $\sigma_i^2 = s_i^2$ and $w_i = s_i^{-2}$, i.e. under the DerSimonian–Laird scenario, equation (9) becomes

$$\tilde{\sigma}^2_{\mathrm{U}} = \frac{T_1^2}{T_1^2 T_2 - 2T_1 T_3 + T_2^2} \left\{ \sum_i \frac{(x_i - \tilde{x}_{\mathrm{GD}})^2}{s_i^4} + \frac{T_2}{T_1} - T_1 \right\}$$
$$= \frac{1}{S_2 - 2S_3 + S_2^2} \left\{ \sum_i (x_i - \tilde{x}_{\mathrm{GD}})^2 \omega_i^2 - \frac{1 - S_2}{T_1} \right\}. \qquad (10)$$

Here $\omega_i = s_i^{-2} / \Sigma s_k^{-2}$, $S_m = \Sigma \omega_i^m$, $m = 2, 3$, and $\tilde{x}_{\mathrm{GD}} = \Sigma \omega_i x_i$ is the Graybill–Deal estimator. It follows that $\tilde{\sigma}^2_{\mathrm{U}}$ has smaller variance than $\tilde{\sigma}^2_{\mathrm{DL}}$ or $\tilde{\sigma}^2_{\mathrm{H}}$ at $\sigma^2 = 0$, strictly so if $p \geqslant 3$.

Because of equation (7), this variance as a function of σ^2 can be written as

$$E(\tilde{\sigma}^2_{\mathrm{U}} - \sigma^2)^2 = 2(T_1^2 T_2 - 2T_1 T_3 + T_2^2)^{-2} \{\sigma^4 (T_1^4 T_4 - 4T_1^3 T_5 + 4T_1^2 T_2 T_4 + 2T_1^2 T_3^2 - 4T_1 T_2^2 T_3 + T_2^4) + 2\sigma^2 T_1 (T_1^3 T_3 - 3T_1^2 T_4 + 3T_1 T_2 T_3 - T_2^3) + T_1^2 T_2 - 2T_1 T_3 + T_2^2\}.$$

At $\sigma^2 = 0$, the mean-squared error, $2(T_1^2 T_2 - 2T_1 T_3 + T_2^2)^{-1}$, coincides with the lower information bound based on the restricted likelihood function, so $\tilde{\sigma}^2_{\mathrm{U}}$ is optimal at this point within the class of all unbiased estimators which depend only on $x_i - x_j$ and s_i^2.

Perhaps the most surprising implication of formulae (4), (5) or (7) is that all unbiased estimators including $\tilde{\sigma}^2_{\mathrm{DL}}$, $\tilde{\sigma}^2_{\mathrm{H}}$ and $\tilde{\sigma}^2_{\mathrm{U}}$ poorly estimate σ^2 when p is small. For $p \leqslant 3$, according to inequality (6), $\mathrm{var}(\tilde{\sigma}^2) > 2\sigma^4/(p-1) \geqslant \sigma^4$, which means that the trivial zero estimator has a smaller mean-squared error than $\tilde{\sigma}^2_{\mathrm{H}}$ or $\tilde{\sigma}^2_{\mathrm{U}}$.

4. Locally optimal estimators

Because of equations (24) and (26) in Appendix A.1 the mean-squared error of $\tilde{\sigma}^2$ can be written as

$$E(\tilde{\sigma}^2 - \sigma^2)^2 = 2\sum_i c_i^2 (\sigma^2 + \sigma_i^2)^2 + 2\sum_{i \neq j} c_{ij}^2 (\sigma^2 + \sigma_i^2)(\sigma^2 + \sigma_j^2) + 2\sum_i \frac{d_i^2 \sigma_i^4}{\nu_i}$$
$$+ \left\{ (c-1)\sigma^2 + \sum_i (c_i + d_i)\sigma_i^2 \right\}^2 = \mathbf{f}^{\mathrm{T}} K \mathbf{f} + (g^{\mathrm{T}} \mathbf{f} - \sigma^2)^2, \qquad (11)$$

which is a quadratic form in the $p(p+1)/2$-dimensional vector $\mathbf{f}$ formed by co-ordinates $c_{ij}, i < j$, and d_i. Here $g^{\mathrm{T}} = (g_1, g_2)^{\mathrm{T}}$, where g_1 is the vector with co-ordinates $2\sigma^2 + \sigma_i^2 + \sigma_j^2$, and g_2 has co-ordinates $E(s_i^2) = \sigma_i^2$. The matrix K is block diagonal, composed of two submatrices K_1 and K_2, where K_2 is diagonal with entries $2\sigma_i^4/\nu_i, i = 1, \ldots, p$, and K_1 is defined in Section 3.

The vector $\mathbf{f}$ minimizing the quadratic form in equation (11) for fixed $\sigma^2 > 0$ and $\sigma_1^2, \ldots, \sigma_p^2$ has the form

$$\mathbf{f} = \sigma^2 (K + g g^{\mathrm{T}})^{-1} g = \frac{\sigma^2 K^{-1} g}{1 + g^{\mathrm{T}} K^{-1} g}. \qquad (12)$$

Let t_1 have co-ordinates

$$0.5(\sigma^2 + \sigma_i^2)^{-1} (\sigma^2 + \sigma_j^2)^{-1} \left\{ \sum_k (\sigma^2 + \sigma_k^2)^{-1} \right\}^{-1}.$$

It is proven in Appendix A.3 that the vector $K_1 t_1$ has co-ordinates $2\sigma^2 + \sigma_i^2 + \sigma_j^2$, i.e. $K_1^{-1} g_1 = t_1$. Thus $\rho_1 = g_1^{\mathrm{T}} K_1^{-1} g_1 = (p-1)/2$, and $2K_2^{-1} g_2 = (\nu_1/\sigma_1^2, \ldots, \nu_p/\sigma_p^2)^{\mathrm{T}}$ and $\rho_2 = g_2^{\mathrm{T}} K_2^{-1} g_2 = \Sigma \nu_k / 2$, so that $g^{\mathrm{T}}K^{-1}g=\rho_1+\rho_2=(\Sigma n_k-1)/2$. Therefore, for $\mathbf{f}$ in equation (12),

$$c_{ij}=\frac{\sigma^2}{\left(\sum_k n_k+1\right)(\sigma^2+\sigma_i^2)(\sigma^2+\sigma_j^2)\sum_k(\sigma^2+\sigma_k^2)^{-1}},$$

$$d_i=\frac{\sigma^2\nu_i}{\left(\sum_k n_k+1\right)\sigma_i^2}.$$

Since $c_{ij}>0$ and $d_i>0$, this locally optimal in class (2) estimator cannot be unbiased.

In the next section we consider the Bayes estimation of σ^2 which requires a prior distribution of this parameter. If only the prior mean, say, β can be specified, we may be interested in an estimator (2) with the smallest mean-squared error at β. With σ_i^2 replaced by s_i^2, equation (12) shows that such an estimator is

$$\tilde{\sigma}^2_{\mathrm{L0}}=\frac{\beta\left\{\sum(x_i-\tilde{x})^2/(\beta+s_i^2)+\sum n_i-p\right\}}{\sum n_i+1}, \tag{13}$$

where $\tilde{x}=\Sigma_i x_i(\beta+s_i^2)^{-1}/\Sigma_i(\beta+s_i^2)^{-1}$. An argument in the next section is made to take β a multiple of Σs_i^2.

5. Bayes estimators

In this section Bayes estimators of σ^2 within class (2) are determined under assumption that σ_i^2 and σ^2 are random independent parameters. Usually more information is available about σ_i^2 than about σ^2, so following the empirical Bayes approach we assume that the prior mean of σ_i^2 can be well estimated by s_i^2, and that the corresponding prior variance v_i is negligible. In contrast, to make the prior distribution Π of σ^2 'non-informative', the prior variance of σ^2, say, v, is supposed to be large, whereas its mean is some β. We have

$$E(\sigma^2 g)=\int\sigma^2 g\,\mathrm{d}\Pi=2v\begin{pmatrix}e\\0\end{pmatrix}+\beta\bar{g},$$

where $\bar{g}^{\mathrm{T}}=(\bar{g}_1,\bar{g}_2)^{\mathrm{T}}$, with $\bar{g}_1$ the vector with co-ordinates $2\beta+s_i^2+s_j^2$, $1\leqslant i<j\leqslant p$, and $\bar{g}_2$ has co-ordinates s_i^2. Thus $\bar{g}_1$ is g_1 from Section 4 where β now stands for σ^2, and s_i^2 for σ_i^2. A similar convention is adapted for matrices K_1 and K_2.

We also put L to be equal to the integrated matrix of second moments of the $p(p+1)/2$-dimensional random vector Y which defines class (2). Thus

$$L=E(YY^{\mathrm{T}})=E\{\mathrm{cov}(Y)\}+E\{E(Y)E(Y^{\mathrm{T}})\},$$

where $E\{\mathrm{cov}(Y)\}$ is a block diagonal matrix which has a form that is similar to that of the matrix K in Section 3. Its first block is the sum of three matrices, $\bar{K}_1$ (with non-zero elements $2(2\beta+s_i^2+s_j^2)^2$ or $2(\beta+s_i^2)^2$), the second is vK_0 with K_0 also of the form K_1 but whose non-zero elements are 8 or 2. The third matrix, which has elements $2(v_i+v_j)$ on the diagonal and non-zero off-diagonal elements $2v_i$, is neglected in the following calculations. The second block of $E\{\mathrm{cov}(Y)\}$ is $\bar{K}_2$, i.e. the diagonal matrix with elements $2(s_i^4+v_i)/\nu_i\approx 2s_i^4/\nu_i$. Up to terms of order v_i,

$$E\{E(Y)E(Y^{\mathrm{T}})\} = 4v\begin{pmatrix} ee^{\mathrm{T}} & 0 \\ 0 & 0 \end{pmatrix} + \bar{g}\bar{g}^{\mathrm{T}},$$

so if $\bar{K}$ denotes the block diagonal matrix formed by $\bar{K}_1$ and $\bar{K}_2$,

$$L = \bar{K} + \bar{g}\bar{g}^{\mathrm{T}} + vRR^{\mathrm{T}}. \tag{14}$$

Here

$$R = \begin{pmatrix} (K_0 + 4ee^{\mathrm{T}})^{1/2} \\ 0 \end{pmatrix}$$

is the matrix of size $p(p+1)/2 \times p$ and rank p.

The Bayes estimator within class (2) is defined by the coefficient vector $L^{-1}E(\sigma^2 g)$. Since the prior variance is supposed to be large, we are interested in the limit

$$\mathbf{f}_{\mathrm{B}} = \lim_{v\to\infty} L^{-1}\left\{2v\begin{pmatrix} e \\ 0 \end{pmatrix} + \beta\bar{g}\right\}.$$

The derivation in Appendix A.3 gives its form

$$\mathbf{f}_{\mathrm{B}} = \begin{pmatrix} \dfrac{e}{p(p+1)} \\ \dfrac{\nu_i\{2\beta - (p-1)\sum s_i^2/p\}}{(p+1)\sum\limits_k (n_k - p + 2)s_i^2} \end{pmatrix}. \tag{15}$$

If $\beta < (p-1)\Sigma s_i^2/p$, the resulting Bayes estimator within class (2),

$$\tilde{\sigma}_{\mathrm{B}}^2 = \frac{\sum(x_i - \bar{x})^2}{p+1} + \frac{\left(\sum n_i - p\right)\{2p\beta - (p-1)\sum s_i^2\}}{\sum(n_i - p + 2)p(p+1)}, \tag{16}$$

takes negative values and must be truncated. Indeed it is Bayes only within the class of linear functions of Y. All such estimators have a positive bias, which can be evaluated from equation (3) when $\tilde{\sigma}_{\mathrm{B}}^2$ is positive.

The factor $(p+1)^{-1}$ in equation (16) is due to the convenient choice of the quadratic loss function which partly was made to accommodate the possibility of $\sigma^2 = 0$. Indeed many other loss functions designed for variance estimation lead to infinite risk functions. However, in our simulations reported in Section 9, we used the absolute error loss. Larger factors, like $(p-1)^{-1}$ or even $(p-3)^{-1}$, $p \geqslant 4$, may be preferable for μ-estimators.

Several versions of the prior mean β can be suggested as default choices. The first is $\beta = 0$ with

$$\tilde{\sigma}_{\mathrm{B0}}^2 = \frac{\sum(x_i - \bar{x})^2}{p+1} - \frac{\left(\sum n_i - p\right)(p-1)\sum s_i^2}{p(p+1)\left(\sum n_i - p + 2\right)}, \tag{17}$$

which requires truncation at zero. Another choice, $\beta = 0.5(p-1)\Sigma s_i^2/p$, leads to a very simple positive estimator,

$$\tilde{\sigma}_{\mathrm{BP}}^2 = \frac{\sum(x_i - \bar{x})^2}{p+1}. \tag{18}$$

In some chemistry applications it is believed that $\beta \approx 3\Sigma s_i^2/p$. Indeed, under our model the sum $\sigma^2 + \Sigma\sigma_i^2/p$ represents the average *reproducibility* error, whereas $\Sigma\sigma_i^2/p$ is the average *repeatability* error. In many homoscedastic analytical chemistry studies the square root of the ratio between these two errors is in the interval from $\frac{1}{2}$ to $\frac{2}{3}$ (Horwitz, 1982; Thompson and Lowthian, 1997). In the author's experience σ^2-estimates rarely if ever exceed $(p-1)\Sigma s_i^2$, in which case the truncated version of equation (16) is to be compared with estimator (13). Highly heterogeneous models for which $\sigma^2 \gg (p-1)\Sigma\sigma_i^2$ or $\beta \gg (p-1)\Sigma s_i^2$ are not very useful in practice. For these reasons optimality for large σ^2 may not be a relevant issue.

The derivation in Appendix A.3 also leads to the form of the Bayes estimator $\tilde{\sigma}^2_{\mathrm{SB}}$ in the class $\Sigma_{1\leqslant i<j\leqslant p} c_{ij}(x_i-x_j)^2+d$, when $\sigma_i^2=s_i^2$, $i=1,\ldots,p$, are assumed to be known. For a diffuse prior distribution with mean β, the Bayes estimator is determined by $c_{ij}\equiv\{p(p+1)\}^{-1}$ and

$$d=\frac{2\beta-(p-1)\sum s_i^2/p}{p+1}.$$

Thus this estimator looks very similar to estimator (16):

$$\tilde{\sigma}^2_{\mathrm{SB}}=\frac{\sum(x_i-\bar{x})^2}{p+1}+\frac{2p\beta-(p-1)\sum s_i^2}{p(p+1)}. \tag{19}$$

When p is large, estimator (19) is numerically close to estimator (17) and to the Hedges estimator (as $\beta/(p+1)$ is small). However, the variance of the latter estimator for $\beta<0.5(p-1)\Sigma s_i^2/p$ is typically larger than the mean-squared error of $\tilde{\sigma}^2_{\mathrm{BP}}$ whose form easily follows from equation (11):

$$E(\tilde{\sigma}^2_{\mathrm{BP}}-\sigma^2)^2=\frac{2\sigma^4(p+1)p^2+2p(p-2)\sum\sigma_i^4+(p^2-2p+3)\left(\sum\sigma_i^2\right)^2}{\{p(p+1)\}^2}.$$

DuMouchel (1990) suggested modelling σ_i^2 as a random multiple γ of given s_i^2 with $\gamma\sim q/\chi^2(q)$. The proof in Appendix A.3 can be extended to cover this case, showing that the Bayes estimator in class (2) is

$$\tilde{\sigma}^2_{\mathrm{DM}}=\frac{\sum(x_i-\bar{x})^2}{p+1}+\frac{2p\beta-(p-1)\sum s_i^2E(\gamma)}{p(p+1)}. \tag{20}$$

There is a body of work on Bayes estimation in meta-analysis especially on hierarchical Bayes modelling (Abrams and Sanso, 1998; Morris and Normand, 1992). Most of the documented procedures do not have an explicit form and rely on asymptotic formulae, and some of them are numerically intensive. The explicit form of the estimators (17), (18) and (19) is one of their advantages. Section 9 suggests that this is not their only advantage.

6. Non-normal distributions

There is considerable interest in alternative, non-normal distributions in meta-analysis (e.g. Lee and Thompson (2008) and Baker and Jackson (2008)). In this section we assume that $\sigma_i^2=s_i^2$ are given, whereas for each i a statistic x_i with the variance $\sigma^2+s_i^2$ has a possibly non-Gaussian distribution such that $E(x_i)=\mu$ and $E(x_i-\mu)^3=0$, and whose fourth moment is a quadratic function of σ^2 and s_i^2, say,

$$E(x_i-\mu)^4=(\kappa+1)(\sigma^2+s_i^2)^2+\xi\sigma^2s_i^2.$$

Here ξ and κ are two constants such that

$$\mathrm{var}\{(x_i-\mu)^2\}=\kappa(\sigma^2+s_i^2)^2+\xi\sigma^2s_i^2.$$

In the Gaussian case $\kappa=2$ and $\xi=0$.

Under this scenario the covariance matrix of the random vector with co-ordinates $(x_i - x_j)^2$, $i < j$, has non-zero entries

$$(1+\kappa/2)(2\sigma^2 + s_i^2 + s_j^2)^2 + \xi(s_i^2 + s_j^2)\{\sigma^2 + (s_i^2 + s_j^2)/2\} + \{(\kappa - \xi)/2 - 1\}(s_i^2 - s_j^2)^2$$

(on the diagonal) and $\kappa(\sigma^2 + s_i^2)^2 + \xi\sigma^2 s_i^2$ corresponding to the pairs of indices with one common index i. Its expectation with regard to the prior distribution of σ^2, $v\bar{K}_0 + \bar{K}_1$, has the same form as in Section 5. Here $\bar{K}_0$ is of the same structure as K_0 and has non-zero elements $2\kappa + 4$ and κ, whereas $\bar{K}_1$ has diagonal elements

$$(1+\kappa/2)(2\beta + s_i^2 + s_j^2)^2 + \xi(s_i^2 + s_j^2)\{\beta + (s_i^2 + s_j^2)/2\} + \{(\kappa - \xi)/2 - 1\}(s_i^2 - s_j^2)^2,$$

and non-zero off-diagonal elements $\kappa(\beta + s_i^2)^2 + \xi\beta s_i^2$.

As in Section 5, given a diffuse prior distribution for σ^2 with some mean β, the Bayes estimator in the class, $\Sigma_{1 \leqslant i < j \leqslant p} c_{ij}(x_i - x_j)^2 + d$, is sought. In this situation, the matrix $L = E(YY^{\mathrm{T}})$ can be written as equation (14) with scalar $\bar{K}_2 = 0$ and $\bar{g}_2 = 1$. The $(p(p-1)/2 + 1) \times p$ matrix R has the same form with K_0 replaced by $\bar{K}_0$.

The calculations in Appendix A.3 show that the Bayes coefficients are $c_{ij} \equiv \{(p+\kappa)(p-1) + 2\}^{-1}$ and

$$d = \{(\kappa p - \kappa + 2)\beta - (p-1)\sum s_i^2\}\{(p+\kappa)(p-1) + 2\}^{-1}.$$

The non-parametric estimator

$$\tilde{\sigma}^2_{\mathrm{NB}} = \frac{p\sum(x_i - \bar{x})^2 + (\kappa p - \kappa + 2)\beta - (p-1)\sum s_i^2}{(p+\kappa)(p-1) + 2} \tag{21}$$

extends estimator (19). For $\kappa > 2$, the x's sample variance component of $\tilde{\sigma}^2_{\mathrm{NB}}$ is shrunk. The limit of estimator (21) for heavy-tailed distributions when $\kappa \to \infty$ is merely the prior mean β.

As an example, assume that x_i have the distribution of the sum of two independent Laplace random variables: one with mean μ and variance s_i^2; another (the study effect) with zero mean and variance σ^2. Then $\kappa = 5$ and $\xi = -6$, and estimator (21) can be compared with the estimator

$$\hat{\sigma} = \sqrt{2}\sum |x_i - \mathrm{median}(x_i)|/(p-1)$$

used in this setting by Rukhin and Possolo (2011), equation (9).

Simulations show that estimator (21) is competitive especially for small σ^2. A similar state of affairs happens in estimation of μ: the weighted median as an estimator of μ is better than the weighted means statistic based on estimator (21) for large σ^2. For smaller σ^2 the situation is reversed.

7. The case when $p = 2$

Many metrology applications of meta-analysis involve comparison of merely two methods. Then $p = 2$, and the (unique in this case) unbiased estimator of σ^2 is

$$\tilde{\sigma}^2_{\mathrm{DL}} = \tilde{\sigma}^2_{\mathrm{H}} = \tilde{\sigma}^2_{\mathrm{U}} = \frac{(x_1 - x_2)^2 - s_1^2 - s_2^2}{2}.$$

This is also the restricted maximum likelihood estimator. Direct calculation or formulae (4) and (5) with $c_{12}=\frac{1}{2}=c_1=c_2=-d_1=-d_2$ show that, for fixed s_1^2 and s_2^2, the variance of this unbiased estimator is

$$\kappa_0^2=\frac{(2\sigma^2+\sigma_1^2+\sigma_2^2)^2}{2}.$$

The quadratic risk of $\tilde{\sigma}_{\mathrm{U}}^2$ has the form

$$\mathrm{var}(\tilde{\sigma}_{\mathrm{U}}^2)=\frac{1}{4}\left\{2(2\sigma^2+\sigma_1^2+\sigma_2^2)^2+\frac{2\sigma_1^4}{\nu_1}+\frac{2\sigma_2^4}{\nu_2}\right\}=\kappa_0^2+\frac{\sigma_1^4}{2\nu_1}+\frac{\sigma_2^4}{2\nu_2}.$$

The Bayes estimator (16) for the prior in Section 5 is

$$\tilde{\sigma}_{\mathrm{B}}^2=\frac{(x_1-x_2)^2}{6}+\frac{(n_1+n_2-2)(4\beta-s_1^2-s_2^2)}{6(n_1+n_2)}.$$

The choices of the prior mean β discussed there when $\beta=(s_1^2+s_2^2)/4$ give a positive estimator,

$$\tilde{\sigma}_{\mathrm{BP}}^2=\frac{(x_1-x_2)^2}{6},$$

and a rule which requires truncation at zero,

$$\tilde{\sigma}_{\mathrm{B0}}^2=\frac{(x_1-x_2)^2}{6}-\frac{(n_1+n_2-2)(s_1^2+s_2^2)}{6(n_1+n_2)},$$

when $\beta=0$. If $\beta=(s_1^2+s_2^2)/4$ estimator (13) takes the form

$$\tilde{\sigma}_{\mathrm{L0}}^2=\frac{(x_1-x_2)^2}{6(n_1+n_2+1)}+\frac{(n_1+n_2-2)(s_1^2+s_2^2)}{4(n_1+n_2+1)}.$$

The form of the Bayes estimator when $\sigma_1^2=s_1^2$ and $\sigma_2^2=s_2^2$ are known can be given for any prior mean β and any prior variance v:

$$\tilde{\sigma}_{\mathrm{SB}}^2=v\frac{(x_1-x_2)^2+4\beta-s_1^2-s_2^2}{6v+(2\beta+s_1^2+s_2^2)^2}+\frac{\beta(2\beta+s_1^2+s_2^2)^2}{6v+(2\beta+s_1^2+s_2^2)^2}.$$

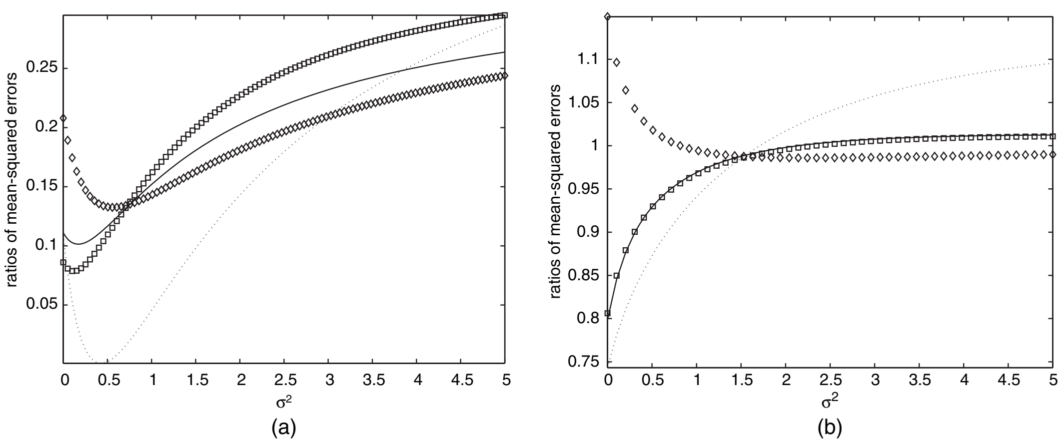

Fig. 1. Plots of ratios of mean-squared errors of estimators $\tilde{\sigma}_{\mathrm{L0}}^2$ (· - - - ·), $\tilde{\sigma}_{\mathrm{SB}}^2$ (———), $\tilde{\sigma}_{\mathrm{B0}}^2$ (□) and $\tilde{\sigma}_{\mathrm{BP}}^2$ (◇) to (a) the mean-squared error of $\tilde{\sigma}_{\mathrm{DL}}^2$ and of corresponding ratios of weighted means estimators of μ to (b) the mean-squared error of $\tilde{x}_{\mathrm{DL}}$

Thus $\tilde{\sigma}^2_{\mathrm{SB}}$ is a convex combination of the Bayes estimator equal to β, when $v=0$, and the Bayes estimator (19) corresponding to $v=\infty$. This interpretation is quite illuminating for any p.

All estimators $\tilde{\sigma}^2_{\mathrm{BP}}$ and truncated versions of $\tilde{\sigma}^2_{\mathrm{SB}}$, $\tilde{\sigma}^2_{\mathrm{L0}}$ and $\tilde{\sigma}^2_{\mathrm{B0}}$ have uniformly smaller quadratic risk than the positive part of $\tilde{\sigma}^2_{\mathrm{U}}$ in the range considered, $0\leqslant\sigma^2\leqslant 5$. Fig. 1 shows the graphs of the ratios of the mean-squared errors $E(\tilde{\sigma}^2-\sigma^2)^2$ for $\tilde{\sigma}^2_{\mathrm{BP}}$, $\tilde{\sigma}^2_{\mathrm{L0}}$ with $\beta=(s_1^2+s_2^2)/4$, and the truncated versions of $\tilde{\sigma}^2_{\mathrm{B0}}$, and $\tilde{\sigma}^2_{\mathrm{SB}}$ with $\beta=0$, to that of $\max(0,\tilde{\sigma}^2_{\mathrm{U}})$ when $n_1=5$ and $n_2=13$, with $\sigma_1^2=s_1^2=0.2$ and $\sigma_2^2=s_2^2=1.8$. These mean-squared errors (as well as the following errors of μ-estimators) were obtained via numerical integration (see Rukhin (2012)).

The dominance of positive σ^2-estimators over $\max(0,\tilde{\sigma}^2_{\mathrm{U}})$ does not really translate into the dominance of μ-estimators. The ratios of quadratic risk functions of the weighted means $\tilde{x}$ for the σ^2-estimators considered to that of $\tilde{x}^2_{\mathrm{DL}}$ are given in the Fig. 1(b). Whereas the locally optimal estimator $\tilde{\sigma}^2_{\mathrm{L0}}$ is considerably better than the positive part of $\tilde{\sigma}^2_{\mathrm{U}}$, the behaviour of the corresponding weighted mean is poor unless σ^2 is small. For this reason estimator (13) will not be considered further. The Bayes estimators lead to better confidence intervals for μ which are discussed in the next section.

8. Confidence intervals for μ

Despite importance of confidence bounds for σ^2, we focus now on a more urgent practical issue, namely that of coverage intervals for μ. For a generic positive estimator $\tilde{\sigma}^2$, let

$$\tilde{\omega}_i=(\tilde{\sigma}^2+s_i^2)^{-1}/\sum_k(\tilde{\sigma}^2+s_k^2)^{-1}$$

be the normalized weights. We employ confidence intervals based on the estimated variance of $\tilde{x}=\Sigma_i\,\tilde{\omega}_i x_i$. One such estimator,

$$\widehat{\mathrm{var}}(\tilde{x})=\left(\sum_i\frac{1}{\tilde{\sigma}^2+s_i^2}\right)^{-1},$$

is a plug-in version of the formula for the variance of $\tilde{\mu}$ from Section 1,

$$\mathrm{var}(\tilde{\mu})=\left(\sum_i\frac{1}{\sigma^2+\sigma_i^2}\right)^{-1}.$$

However, it is known to underestimate the true variance and commonly leads to intervals which are too narrow (Brockwell and Gordon, 2001; Rukhin, 2009).

A better estimator is based on the procedure of Horn *et al.* (1975), $\Sigma_i\,\tilde{\omega}_i^2(1-\tilde{\omega}_i)^{-1}(x_i-\tilde{x})^2$. Another corresponds to the weighted sample variance (the so-called external consistency estimator of the variance), $\Sigma_i\,\tilde{\omega}_i(x_i-\tilde{x})^2/(p-1)$.

Under the assumption of the (approximate) t-distribution of $(\tilde{x}-\mu)/\sqrt{\widehat{\mathrm{var}}(\tilde{x})}$, the resulting intervals have the form

$$\tilde{x}\pm t_{1-\alpha/2}(p-1)\sqrt{\left\{\sum_i\frac{\tilde{\omega}_i^2(x_i-\tilde{x})^2}{1-\tilde{\omega}_i}\right\}},\qquad(22)$$

and

$$\tilde{x}\pm t_{1-\alpha/2}(p-1)\sqrt{\left\{\sum_i\frac{\tilde{\omega}_i(x_i-\tilde{x})^2}{p-1}\right\}}.\qquad(23)$$

The common methods of standard error evaluation for the treatment effect μ do not explicitly take into account the sample sizes n_i, but the standard errors of confidence intervals based on estimator (17) depend on the degrees of freedom. However, potential gains of using estimator (17) rather than estimator (19) with $\beta=0$ can be noticeable only for small p and n_i and large σ^2.

9. Simulation results

We report here some results of the numerical comparison of Bayes estimators $\tilde{\sigma}^2_{\mathrm{SB}}$, $\tilde{\sigma}^2_{\mathrm{BP}}$ and $\tilde{\sigma}^2_{\mathrm{B0}}$, with unbiased estimators $\tilde{\sigma}^2_{\mathrm{DL}}$, $\tilde{\sigma}^2_{\mathrm{H}}$ and $\tilde{\sigma}^2_{\mathrm{U}}$ for $p=3$ and $p=10$. More precisely, the positive parts of all these estimators are considered. We also evaluated characteristics of $\tilde{\sigma}^2_{\mathrm{MP}}$ mentioned in Section 2. This comparison was performed with regard to the confidence intervals (22) and (23) based on these estimators, in terms of their mean absolute errors, $E|\tilde{\sigma}^2-\sigma^2|$, and for the mean-squared errors of the corresponding μ-estimators, $E(\tilde{x}-\mu)^2$. In the Monte Carlo simulation we used randomly chosen sample sizes n_i with the uniform distribution over integers from 4 to 12 with 25000 runs for each of the values of $\sigma^2=0, 0.1, \ldots, 2$. The error variances σ_i^2 were taken to have an inverted χ^2-distribution $(\nu_i-2)/\chi^2(\nu_i)$, so that $E(\sigma_i^2)=1$.

Fig. 2 displays the coverage probability of these intervals reported as a function of σ^2 for a nominal 95% confidence coefficient. The intervals (22) (and (23) not shown here as intervals (22) outperformed intervals (23) in our simulations) based on the DerSimonian–Laird estimator, $\tilde{\sigma}^2_{\mathrm{U}}$ or on the Hedges estimator have lower than the stated confidence level. The confidence intervals based on $\tilde{\sigma}^2_{\mathrm{SB}}$ and $\tilde{\sigma}^2_{\mathrm{B0}}$ maintain the nominal confidence coefficient much better. However by far the best in this regard is the interval which uses $\tilde{\sigma}^2_{\mathrm{BP}}$ and which sustains the stated confidence almost perfectly. The q–q-plot of $(\tilde{x}-\mu)/\sqrt{\tilde{\mathrm{var}}(\tilde{x})}$ for the estimator $\tilde{x}$ based on $\tilde{\sigma}^2_{\mathrm{BP}}$ against percentiles of the $t(p-1)$-distribution exhibits a fairly straight line. For larger values of p ($p=20, 30, 50$), the two intervals (22) and (23) were compared by Sidik and Jonkman (2007) who also recommended intervals (22).

Fig. 3 shows that the mean-squared error of the μ-estimator based on $\tilde{\sigma}^2_{\mathrm{BP}}$ is one of the smallest among all methods considered. When p exceeds 10, the behaviour of the unbiased variance estimator $\tilde{\sigma}^2_{\mathrm{U}}$ and $\tilde{\sigma}^2_{\mathrm{MP}}$ resembles that of the DerSimonian–Laird statistic. Then the estimators $\tilde{\sigma}^2_{\mathrm{SB}}$, $\tilde{\sigma}^2_{\mathrm{B0}}$ and $\tilde{\sigma}^2_{\mathrm{H}}$ are very similar and inferior to $\tilde{\sigma}^2_{\mathrm{DL}}$ and to $\tilde{\sigma}^2_{\mathrm{U}}$ for small σ^2, but outperform them for larger σ^2.

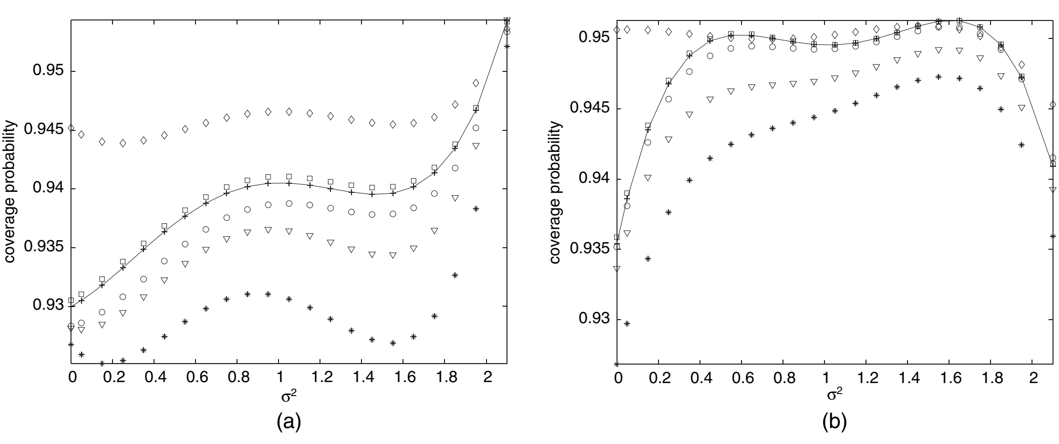

Fig. 2. Plots of coverage intervals (22) based on σ^2-estimators $\tilde{\sigma}^2_{\mathrm{SB}}$ (———), $\tilde{\sigma}^2_{\mathrm{B0}}$ (□), $\tilde{\sigma}^2_{\mathrm{BP}}$ (◇), $\tilde{\sigma}^2_{\mathrm{U}}$ ($*$), $\tilde{\sigma}^2_{\mathrm{H}}$ (+), $\tilde{\sigma}^2_{\mathrm{DL}}$ (∇) and $\tilde{\sigma}^2_{\mathrm{MP}}$ (○) when (a) $p=3$, and of the same intervals when (b) $p=10$

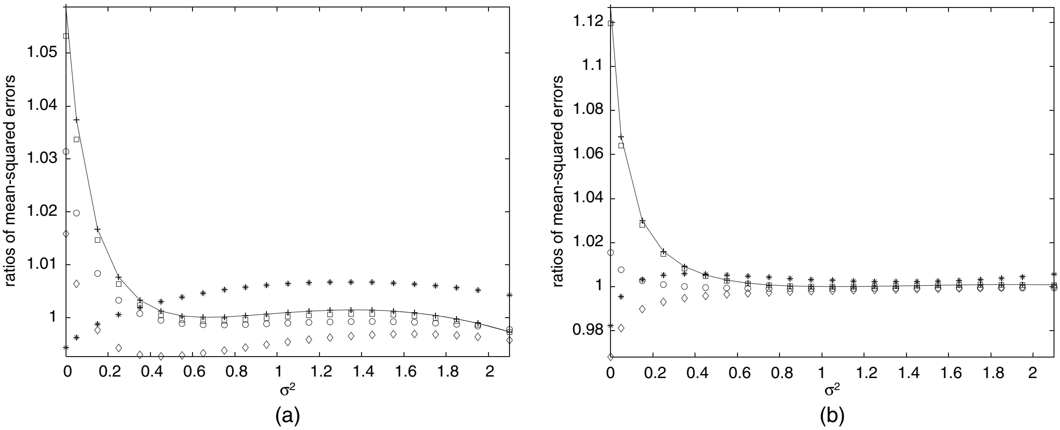

Fig. 3. Plots of ratios of quadratic risk functions of μ-estimators based on $\tilde{\sigma}^2_{\mathrm{SB}}$ (———), $\tilde{\sigma}^2_{\mathrm{B0}}$ ($\square$), $\tilde{\sigma}^2_{\mathrm{BP}}$ ($\diamondsuit$), $\tilde{\sigma}^2_{\mathrm{U}}$ ($*$), $\tilde{\sigma}^2_{\mathrm{H}}$ ($+$) and $\tilde{\sigma}^2_{\mathrm{MP}}$ ($\bigcirc$) to the mean-squared error of $\tilde{x}_{\mathrm{DL}}$ when (a) $p=3$, and of the same ratios when (b) $p=10$

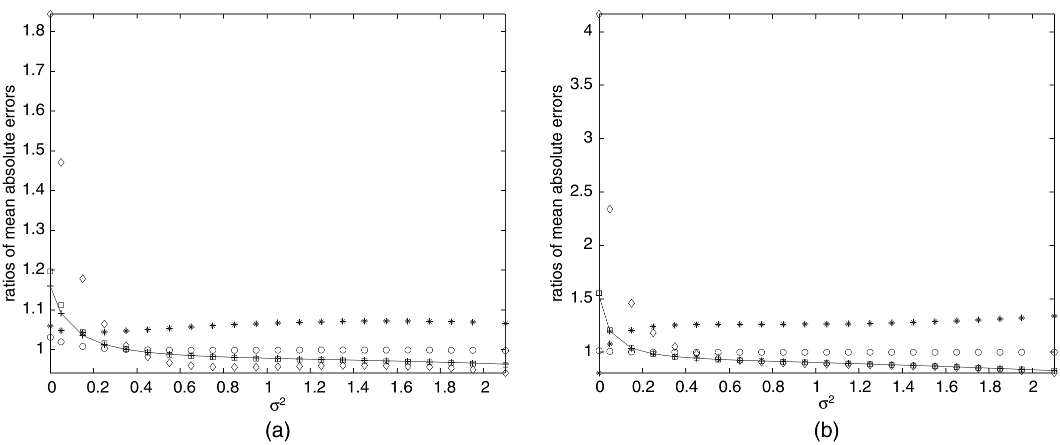

Fig. 4. Plots of ratios of absolute errors functions of σ^2-estimators $\tilde{\sigma}^2_{\mathrm{SB}}$ (———), $\tilde{\sigma}^2_{\mathrm{B0}}$ ($\square$), $\tilde{\sigma}^2_{\mathrm{BP}}$ ($\diamondsuit$), $\tilde{\sigma}^2_{\mathrm{U}}$ ($*$), $\tilde{\sigma}^2_{\mathrm{H}}$ ($+$) and $\tilde{\sigma}^2_{\mathrm{MP}}$ ($\bigcirc$) to the absolute error of $\tilde{\sigma}^2_{\mathrm{DL}}$ when (a) $p=3$, and of the same ratios when (b) $p=10$

The Bayes estimators are better than the traditional estimators in terms of the mean absolute error, except that, for larger p, $\tilde{\sigma}^2_{\mathrm{BP}}$ is losing to other estimators for small values of σ^2, as is seen from Fig. 4.

10. Example

To facilitate global trade, Le Comite International des Poids et Mesures initiated a series of international interlaboratory studies which are designed to establish the degree of equivalence between participating national metrology institutes. One of such studies is international fluid flow comparisons of the air speed measurement CCM.FF-K3; Terao *et al.* (2007). An ultrasonic anemometer chosen as a transfer standard was circulated between four national metrology institutes who reported calibration results at 2 m s^{-1} and at 20 m s^{-1}. Thus x_i given in the tables represents the ratio of the laboratory's reference air speed to the speed measured by the transfer standard, and as such is a dimensionless quantity.

Table 1. Air speed data for four institutes (2 m s^{-1})

x_i	0.9993	1.0090	1.0196	1.0270
s_i	0.0026	0.0027	0.0017	0.0025

The data for 2 m s^{-1} are quite scattered and the individual coverage intervals for μ overlapped only pairwise, so it was decided to use the median as an estimator of μ. Table 1 gives these data along with associated uncertainties.

All σ^2-estimators considered are positive and different,

$$\begin{array}{cccccc} \tilde{\sigma}^2_{\mathrm{DL}} & \tilde{\sigma}^2_{\mathrm{U}} & \tilde{\sigma}^2_{\mathrm{H}} & \tilde{\sigma}^2_{\mathrm{MP}} & \tilde{\sigma}^2_{\mathrm{BP}} & \tilde{\sigma}^2_{\mathrm{SB}} \\ 0.1626\times 10^{-3} & 0.1384\times 10^{-3} & 0.1832\times 10^{-3} & 0.1827\times 10^{-3} & 0.1134\times 10^{-3} & 0.1116\times 10^{-3} \end{array}$$

which suggests that a random-effects model may be approprate. The smallest values correspond to the Bayes estimators $\tilde{\sigma}^2_{\mathrm{BP}}$ and $\tilde{\sigma}^2_{\mathrm{SB}}$, but all of them have the same order as $(p-1)\Sigma s_i^2/p = 0.1739\times 10^{-3}$. The sample sizes are not available so $\tilde{\sigma}^2_{\mathrm{B0}}$ cannot be used.

The μ-estimators evaluated according to the various methods above turn out to be quite close to 1.01176, and the coverage intervals (22) do not differ very much despite a fairly large degree of heterogeneity. The usual interval, $\tilde{x}_{\mathrm{GD}} \pm t_{1-\alpha/2}(p-1)\sqrt{\widehat{\mathrm{var}}(\tilde{x}_{\mathrm{GD}})} = 1.0141 \pm 0.0035$, with $\widehat{\mathrm{var}}(\tilde{x}_{\mathrm{GD}})$ as in Section 8, seems to be too short.

All our μ-values are closer to 1 than the reference value $\mu_{\mathrm{ref}} = 1.0143$ suggested in the final report for CCM.FF-K3. The final uncertainty, 0.0077, reported there is considerably smaller than the typical value of the half-width 95% confidence intervals, which is about 0.0218.

The results obtained at the speed 20 m s^{-1} are given in Table 2.

Although these data were deemed to be considerably less heterogeneous, all our estimators of σ^2 are again positive (although an order of magnitude smaller than those for 2 m s^{-1}):

$$\begin{array}{cccccc} \tilde{\sigma}^2_{\mathrm{DL}} & \tilde{\sigma}^2_{\mathrm{U}} & \tilde{\sigma}^2_{\mathrm{H}} & \tilde{\sigma}^2_{\mathrm{MP}} & \tilde{\sigma}^2_{\mathrm{BP}} & \tilde{\sigma}^2_{\mathrm{SB}} \\ 0.2008\times 10^{-4} & 0.1165\times 10^{-4} & 0.2568\times 10^{-4} & 0.2576\times 10^{-4} & 0.1854\times 10^{-4} & 0.2167\times 10^{-4} \end{array}$$

The μ-estimators

$$\begin{array}{cccccc} \tilde{x}_{\mathrm{DL}} & \tilde{x}_{\mathrm{U}} & \tilde{x}_{\mathrm{H}} & \tilde{x}_{\mathrm{MP}} & \tilde{x}_{\mathrm{BP}} & \tilde{x}_{\mathrm{SB}} \\ 1.0087 & 1.0090 & 1.0087 & 1.0087 & 1.0088 & 1.0087 \end{array}$$

are more dispersed than in the previous case but are all closer to 1 than the reference value given in Terao *et al.* (2007) as 1.0113. The uncertainties, $\sqrt{\widehat{\mathrm{var}}(\tilde{x})}$, estimated after intervals (22), are

$$\begin{array}{cccccc} \tilde{x}_{\mathrm{DL}} & \tilde{x}_{\mathrm{U}} & \tilde{x}_{\mathrm{H}} & \tilde{x}_{\mathrm{MP}} & \tilde{x}_{\mathrm{BP}} & \tilde{x}_{\mathrm{SB}} \\ 0.0087 & 0.0086 & 0.0088 & 0.0088 & 0.0087 & 0.0088 \end{array}$$

and these values can be contrasted with a much smaller reported uncertainty of 0.0012.

This is just one out of many interlaboratory comparisons whose analysis could have benefited from the random-effects model.

Table 2. Air speed data for four institutes (20 m s^{-1})

x_i	1.0064	1.0080	1.0128	1.0120
s_i	0.0026	0.0030	0.0014	0.0015

11. Conclusions

The Bayes estimators (17) and (18) show promise for meta-analysis in the case of small to moderate numbers of studies and can be seriously considered in these applications. Better knowledge of the prior mean may lead to further improvement in the performance of estimator (19).

A meta-analyst must be willing to use different estimates of the between-study variance σ^2 for different purposes: one to minimize the variance of the treatment effect statistic; another to construct a reliable confidence interval for this parameter; yet another to estimate σ^2 itself! It may be practically impossible to find a much better point estimator of μ than the DerSimonian–Laird statistic, but $\tilde{\sigma}^2_{\mathrm{DL}}$ is not appropriate if the heterogeneity variance itself is of interest, or when a confidence interval for μ is needed. For the latter goal $\tilde{\sigma}^2_{\mathrm{BP}}$ can be highly recommended; for the former $\tilde{\sigma}^2_{\mathrm{B0}}$ or $\tilde{\sigma}^2_{\mathrm{SB}}$ are good candidates.

Appendix A

A.1. Bias and variance of $\tilde{\sigma}^2$

Formula (2) shows that

$$E(\tilde{\sigma}^2|s_1^2,\ldots,s_p^2)=\sum_{i<j} c_{ij}(2\sigma^2+\sigma_i^2+\sigma_j^2)+\sum d_i s_i^2 \qquad (24)$$

$$=\sigma^2\sum_{i,j} c_{ij}+\sum c_i\sigma_i^2+\sum d_i s_i^2,$$

$$E\{\tilde{\sigma}^2(s_i^2-\sigma_i^2)\}=2d_i\sigma_i^4/\nu_i \qquad (25)$$

and

$$\mathrm{var}\{E(\tilde{\sigma}^2|s_1^2,\ldots,s_p^2)\}=\mathrm{var}\Big(\sum d_i s_i^2\Big)=\sum\frac{2d_i^2\sigma_i^4}{\nu_i}. \qquad (26)$$

Since

$$\mathrm{cov}\{(x_i-x_j)^2,(x_k-x_l)^2\}=\begin{cases}2(2\sigma^2+\sigma_i^2+\sigma_j^2)^2 & (i,j)=(k,l),\\ 2(\sigma^2+\sigma_i^2)^2 & (i,j),(k,l)\text{ share index } i,\\ 0 & \text{otherwise}\end{cases}$$

we have

$$\mathrm{var}(\tilde{\sigma}^2|s_1^2,\ldots,s_p^2)=\mathrm{var}\Big\{\sum_{i<j} c_{ij}(x_i-x_j)^2\Big\}$$

$$=\sum_{i<j}\sum_{k<l} c_{ij}c_{kl}\,\mathrm{cov}\{(x_i-x_j)^2,(x_k-x_l)^2\}$$

$$=2\sum_{1\leqslant i<j\leqslant p} c_{ij}^2(2\sigma^2+\sigma_i^2+\sigma_j^2)^2+2\sum_{i\neq j\neq k} c_{ij}c_{ik}(\sigma^2+\sigma_i^2)^2$$

$$=2\sum_i c_i^2(\sigma^2+\sigma_i^2)^2+2\sum_{i\neq j} c_{ij}^2(\sigma^2+\sigma_i^2)(\sigma^2+\sigma_j^2),$$

which coincides with equation (7).

The formula for the variance of $\tilde{\sigma}^2$ follows as

$$\mathrm{var}(\tilde{\sigma}^2)=E\{\mathrm{var}(\tilde{\sigma}^2|s_1^2,\ldots,s_p^2)\}+\mathrm{var}\{E(\tilde{\sigma}^2|s_1^2,\ldots,s_p^2)\},$$

and this identity gives formula (11) for the mean-squared error.

In the setting of Section 6, if the variance of $(x_i-\mu)^2$ is $(\sigma^2+s_i^2)\kappa+\xi\sigma^2 s_i^2$, then

$$\operatorname{cov}\{(x_i-x_j)^2,(x_k-x_l)^2\}=(1+\kappa/2)(2\sigma^2+s_i^2+s_j^2)^2+\xi(s_i^2+s_j^2)\{\sigma^2+(s_i^2+s_j^2)/2\}\\+\{(\kappa-\xi)/2-1\}(s_i^2-s_j^2)^2,$$

$(i,j)=(k,l)$, or equals $\kappa(\sigma^2+s_i^2)^2+\xi\sigma^2 s_i^2$, (i,j), if (k,l) have one common index i, and is 0, otherwise.

A.2. Proof of inequality (6) and of the formulae for $K_1^{-1}g_1$ and $K_1^{-1}e$

To prove inequality (6), note that, if E_k, $k=0,1,\ldots,p$, denote elementary symmetric polynomials in $w_1,\ldots,w_p$, then $T_1=E_1$, $T_2=E_1^2-2E_2$ and $T_3=E_1^3-3E_1E_2+3E_3$, so

$$\frac{T_1^2T_2-2T_1T_3+T_2^2}{(T_1^2-T_2)^2}=1-\frac{3E_1E_3}{2E_2^2}.$$

Since $E_1E_3\leqslant 2(p-2)E_2^2/\{3(p-1)\}$ (Beckenbach and Bellman (1961), chapter 1, section 12), inequality (6) follows.

To find $K_1^{-1}g$, observe that, for $1\leqslant i<j\leqslant p$, the vector K_1t_1 has co-ordinates,

$$\frac{1}{\sum_k(\sigma^2+\sigma_k^2)^{-1}}\left\{\frac{(2\sigma^2+\sigma_i^2+\sigma_j^2)^2}{(\sigma^2+\sigma_i^2)(\sigma^2+\sigma_j^2)}+\sum_{k:k\neq i,j}\frac{\sigma^2+\sigma_i^2}{\sigma^2+\sigma_k^2}+\sum_{k:k\neq i,j}\frac{\sigma^2+\sigma_j^2}{\sigma^2+\sigma_k^2}\right\},$$

which can be easily seen to coincide with those of g_1.

The desired formula for c_i^0 follows from equation (8) by summing up over j, $j\neq i$. The vector K_1c^0 has (i,j)th co-ordinate equal to

$$2\left\{\left(\frac{1}{w_i}+\frac{1}{w_j}\right)^2c_{ij}^0+\frac{1}{w_i^2}\sum_{k:k\neq i,j}c_{ik}^0+\frac{1}{w_j^2}\sum_{k:k\neq i,j}c_{jk}^0\right\}=2\left(\frac{2c_{ij}^0}{w_iw_j}+\frac{c_i^0}{w_i^2}+\frac{c_j^0}{w_j^2}\right)=\frac{2T_1^2}{T_1^2T_2-2T_1T_3+T_2^2},$$

which does not depend on i and j.

A.3. Proof of equation (15)

Since $K_0e=4pe$, $4ee^{\mathrm{T}}e=2p(p-1)e$,

$$RR^{\mathrm{T}}\begin{pmatrix}e\\0\end{pmatrix}=2p(p+1)\begin{pmatrix}e\\0\end{pmatrix}.$$

We have (Harville (1997), theorem 18.2.8)

$$L^{-1}R=[(\bar K+\bar g\bar g^{\mathrm{T}})^{-1}-v(\bar K+\bar g\bar g^{\mathrm{T}})^{-1}R\{I+vR^{\mathrm{T}}(\bar K+\bar g\bar g^{\mathrm{T}})^{-1}R\}^{-1}R^{\mathrm{T}}(\bar K+\bar g\bar g^{\mathrm{T}})^{-1}]R\\=(\bar K+\bar g\bar g^{\mathrm{T}})^{-1}R\{I+vR^{\mathrm{T}}(\bar K+\bar g\bar g^{\mathrm{T}})^{-1}R\}^{-1},$$

and, since the matrix $R^{\mathrm{T}}(\bar K+\bar g\bar g^{\mathrm{T}})^{-1}R$ is non-singular,

$$\lim_{v\to\infty}vL^{-1}RR^{\mathrm{T}}\begin{pmatrix}e\\0\end{pmatrix}=(\bar K+\bar g\bar g^{\mathrm{T}})^{-1}R\{R^{\mathrm{T}}(\bar K+\bar g\bar g^{\mathrm{T}})^{-1}R\}^{-1}R^{\mathrm{T}}\begin{pmatrix}e\\0\end{pmatrix}.$$

It follows from Section 4 that the vector $\bar K_1^{-1}\bar g_1$ has co-ordinates

$$0.5(\beta+s_i^2)^{-1}(\beta+s_j^2)^{-1}\{\sum_k(\beta+s_k^2)^{-1}\}^{-1},$$

which implies formulae $\rho_1=\bar g_1^{\mathrm{T}}\bar K_1^{-1}\bar g_1=(p-1)/2$ and $e^{\mathrm{T}}\bar g_1=(p-1)\Sigma_k(\beta+s_k^2)$. The co-ordinates of $\bar K_2^{-1}\bar g_2$ are $\nu_i/(2s_i^2)$, so $\rho_2=\bar g_2^{\mathrm{T}}\bar K_2^{-1}\bar g_2=(\Sigma_kn_k-p)/2$. Using the identity

$$(\bar K+\bar g\bar g^{\mathrm{T}})^{-1}=\bar K^{-1}-\frac{\bar K^{-1}\bar g\bar g^{\mathrm{T}}\bar K^{-1}}{1+\rho_1+\rho_2}$$

(Harville (1997), corollary 18.2.10), we obtain

$$R^{\mathrm{T}}(\bar{K}+\bar{g}\bar{g}^{\mathrm{T}})^{-1}R=(K_0+4ee^{\mathrm{T}})^{1/2}\left(\bar{K}_1^{-1}-\frac{\bar{K}_1^{-1}\bar{g}_1\bar{g}_1^{\mathrm{T}}\bar{K}_1^{-1}}{1+\rho_1+\rho_2}\right)(K_0+4ee^{\mathrm{T}})^{1/2},$$

$$\{R^{\mathrm{T}}(\bar{K}+\bar{g}\bar{g}^{\mathrm{T}})^{-1}R\}^{-1}=(K_0+4ee^{\mathrm{T}})^{-1/2}\left(\bar{K}_1+\frac{\bar{g}_1\bar{g}_1^{\mathrm{T}}}{1+\rho_2}\right)(K_0+4ee^{\mathrm{T}})^{-1/2},$$

$$R\{R^{\mathrm{T}}(\bar{K}+\bar{g}\bar{g}^{\mathrm{T}})^{-1}R\}^{-1}R^{\mathrm{T}}=\begin{pmatrix}\bar{K}_1+(1+\rho_2)^{-1}\bar{g}_1\bar{g}_1^{\mathrm{T}} & 0\\ 0 & 0\end{pmatrix},$$

so

$$\lim_{v\to\infty} vL^{-1}RR^{\mathrm{T}}\begin{pmatrix}e\\0\end{pmatrix}=\begin{pmatrix}e\\-(1+\rho_2)^{-1}e^{\mathrm{T}}\bar{g}_1\bar{K}_2^{-1}\bar{g}_2\end{pmatrix}. \tag{27}$$

Since

$$\begin{aligned}\lim_{v\to\infty} L^{-1}\bar{g}&=\lim_{v\to\infty}\frac{(\bar{K}+vRR^{\mathrm{T}})^{-1}\bar{g}}{1+\bar{g}^{\mathrm{T}}(\bar{K}+vRR^{\mathrm{T}})^{-1}\bar{g}^{\mathrm{T}}}\\&=\frac{\{\bar{K}^{-1}-\bar{K}^{-1}R(R^{\mathrm{T}}\bar{K}^{-1}R)^{-1}R^{\mathrm{T}}\bar{K}^{-1}\}\bar{g}}{1+\rho_1+\rho_2-\bar{g}^{\mathrm{T}}\bar{K}^{-1}R(R^{\mathrm{T}}\bar{K}^{-1}R)^{-1}R^{\mathrm{T}}\bar{K}^{-1}\bar{g}}\\&=\begin{pmatrix}0\\(1+\rho_2)^{-1}\bar{K}_2^{-1}\bar{g}_2\end{pmatrix}.\end{aligned}$$

It follows that

$$\mathbf{f}_{\mathrm{B}}=\left(\begin{array}{c}\dfrac{e}{p(p+1)}\\ \dfrac{\nu_i[\beta-(p-1)\{p(p+1)\}^{-1}\sum_k(\beta+s_k^2)]}{\sum_k(n_k-p+2)s_i^2}\end{array}\right),$$

which proves equation (15).

For any non-singular matrix K, the limit $K^{1/2}(K+vRR^{\mathrm{T}})^{-1}K^{1/2}$ for $v\to\infty$ is a projection matrix onto the subspace spanned by the columns of the matrix $K^{-1/2}R$. In our situation $K=\bar{K}+\bar{g}\bar{g}^{\mathrm{T}}$, and the vector $(e^{\mathrm{T}},0)^{\mathrm{T}}$ is orthogonal to this subspace. This fact explains why equation (27) is finite. To obtain a similar result when $\sigma_i^2=s_i^2$ are given, one can put $\bar{K}_2\to 0$ and $\bar{g}_2=1$ in the formulae above, so that $\bar{K}_2^{-1}\bar{g}_2/(1+\rho_2)\to 1$.

In the situation of Section 6, $K_0e=(2p+\kappa-2)e$, and the form of the Bayes coefficients follows.

References

Abrams, K. and Sanso, B. (1998) Approximate Bayes inference for random effects meta-analysis. *Statist. Med.*, **17**, 201–218.

Baker, R. and Jackson, D. (2008) A new approach to outliers in meta-analysis. *Hlth Care Mangmnt Sci.*, **11**, 121–131.

Beckenbach, E. F. and Bellman, R. (1961) *Inequalities*. Berlin: Springer.

Biggerstaff, B. and Tweedie, R. L. (1997) Incorporating variability of estimates of heterogeneity in the random effects model in meta-analysis. *Statist. Med.*, **16**, 753–768.

Böhning, D., Malzahn, U., Dietz, E., Schlattmann, P., Viwatwongkasem, C. and Biggeri, A. (2002) Some general points in estimating heterogeneity variance with the DerSimonian-Laird estimator. *Biostatistics*, **3**, 445–457.

Böhning, D. and Sarol, J. (2000) Estimating risk difference in multicenter studies under baseline-risk heterogeneity. *Biometrics*, **56**, 304–308.

Böhning, D., Sarol, J., Rattanasiri, S., Viwatwongkasem, C. and Biggeri, A. (2004) A comparison of non-iterative and iterative estimators of heterogeneity variance for the standardized mortality ratio. *Biostatistics*, **5**, 61–74.

Brockwell, S. E. and Gordon, I. R. (2001) A comparison of statistical methods for meta-analysis. *Statist. Med.*, **20**, 825–840.

DerSimonian, R. and Kacker, R. (2007) Random-effects model for meta-analysis of clinical trials: an update. *Contemp. Clin. Trials*, **28**, 105–114.

DerSimonian, R. and Laird, N. (1986) Meta-analysis in clinical trials. *Contr. Clin. Trials*, **7**, 177–188.
DuMouchel, W. (1990) Bayesian meta-analysis. In *Statistical Methodology in Pharmaceutical Sciences* (ed. D. Berry), pp. 509–529. New York: Dekker.
Friedman, L. (2000) Estimators of random effects variance components in meta-analysis. *J. Educ. Behav. Statist.*, **25**, 1–12.
Harville, D. A. (1997) *Matrix Algebra from a Statistician's Perspective*. New York: Springer.
Hedges, L. V. (1983) A random-effects model for effect sizes. *Psychol. Bull.*, **93**, 388–395.
Higgins, J. P. T., Thompson, S. G. and Spiegelhalter, D. J. (2009) A re-evaluation of random-effects meta-analysis. *J. R. Statist. Soc.* A, **172**, 137–159.
Horn, S. A., Horn, R. A. and Duncan, D. B. (1975) Estimating heteroscedastic variance in linear models. *J. Am. Statist. Ass.*, **70**, 380–385.
Horwitz, W. (1982) Evaluation of analytical methods used for regulation of foods and drugs. *Analyt. Chem.*, **54**, 67A–76A.
Jackson, D., Bowden, J. and Baker, R. (2010) How does the DerSimonian and Laird procedure for random effects meta-analysis compare with its more efficient but harder to compute counterparts? *J. Statist. Planng Inf.*, **140**, 961–970.
Lee, K. J. and Thompson, S. G. (2008) Flexible parametric models for random effects distributions. *Statist. Med.* **27**, 418–434.
Malzahn, U., Böhning, D. and Holling, H. (2000) Nonparametric estimation of heterogeneity variance for the standardised difference used in meta-analysis. *Biometrika*, **87**, 619–632.
Mandel, J. and Paule, R. C. (1970) Interlaboratory evaluation of a material with unequal number of replicates. *Analyt. Chem.*, **42**, 1194–1197.
Morris, C. N. (1983) Parametric empirical Bayes inference: theory and applications. *J. Am. Statist. Ass.*, **78**, 47–65.
Morris, C. N. and Normand, S. L. (1992) Hierarchical models for combining information and for meta-analyses. In *Bayesian Statistics* 4 (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), pp. 321–344. Oxford: Oxford University Press.
Rukhin, A. L. (2009) Weighted means statistics in interlaboratory studies. *Metrologia*, **46**, 323–331.
Rukhin, A. L. (2012) Confidence intervals for the treatment effect from the restricted maximum likelihood. *J. Statist. Planng Inf.*, **142**, 1999–1208.
Rukhin, A. L. and Possolo, A. (2011) Laplace random effects models for interlaboratory studies. *Computnl Statist. Data Anal.*, **55**, 1815–1827.
Rukhin, A. L. and Vangel, M. G. (1998) Estimation of a common mean and weighted means statistics. *J. Am. Statist. Ass.*, **93**, 303–308.
Sidik, K. and Jonkman, J. (2007) A comparison of heterogeneity variance estimators in combining results of studies. *Statist. Med.*, **26**, 1964–1981.
Terao, Y., van der Beek, M., Yeh, T. T. and Müller, H. (2007) Final report on the CIPM air speed key comparison (CCM.FF-K3). *Metrologia*, **44**, tech. suppl. 07009.
Thompson, M. and Lowthian, P. J. (1997) The Horwitz function revisited. *J. Ass. Off. Analyt. Chem. Int.*, **80**, 676–679.
Viechtbauer, W. (2005) Bias and efficiency of meta-analytic variance estimators in the random-effects model. *J. Educ. Behav. Statist.*, **30**, 261–293.