Research Article

Statistics in Medicine

Received 11 September 2008, Accepted 19 March 2009 Published online 30 April 2009 in Wiley Interscience

(www.interscience.wiley.com) DOI: 10.1002/sim.3602

Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses

Dan Jackson,*† Ian R. White and Simon G. Thompson

Multivariate meta-analysis is increasingly used in medical statistics. In the univariate setting, the non-iterative method proposed by DerSimonian and Laird is a simple and now standard way of performing random effects meta-analyses. We propose a natural and easily implemented multivariate extension of this procedure which is accessible to applied researchers and provides a much less computationally intensive alternative to existing methods. In a simulation study, the proposed procedure performs similarly in almost all ways to the more established iterative restricted maximum likelihood approach. The method is applied to some real data sets and an extension to multivariate meta-regression is described.

Keywords: multivariate meta-analysis; random effects model; moment estimator; meta-regression

1. Introduction

Meta-analysis, the statistical process of combining the results from separate studies concerned with the same treatment or issue, is frequently used in medical statistics and other fields. Typically, relatively simple procedures are used when combining the results, although more sophisticated methodology is also available [1]; for example, meta-regression [2, 3] provides the means to investigate the associations of study-level covariates with the treatment effect and explain between-study heterogeneity.

The focus of this paper is multivariate meta-analysis [4–6], in which studies provide measures of multiple outcomes of interest: for example, details of treatment effect on two or more clinical outcomes. These methods may be extended to provide multivariate meta-regressions; Jackson *et al.* [7] describe a bivariate meta-regression where the mean of both study outcomes depends on time, and also provide a model for the possibility of reporting bias.

Just as in univariate meta-analysis, both fixed and random effects models can be used in the multivariate context. Fixed effects (common underlying effect across all studies) is a strong assumption in the univariate case [8, 9] and this becomes an even stronger assumption in the multivariate setting: it seems generally implausible that there is no between-study heterogeneity in *any* of the various outcomes of interest. Hence, the random effects model will be adopted as the standard model here.

Standard univariate random effects meta-analysis involves two parameters: the mean and the variance of the random effects distribution. Conventionally, the between-study variance is estimated and then is treated as fixed and known when making inferences about the treatment effect [10], a justifiable approximation if there is a sufficiently large number of studies. Typically this variance is estimated using the non-iterative method of moments suggested by DerSimonian and Laird [11], although maximum likelihood [12] and restricted maximum likelihood (REML) estimation [13] are more computationally intensive alternatives that were also examined by DerSimonian and Laird. In the multivariate setting the between-study variance matrix is typically estimated using REML, for example, using SAS Proc Mixed [4]. The standard use of the estimated between-study variance matrix as the true matrix when pooling the studies' results will be adopted, although Kenward and Roger [14] give small sample procedures where REML is used.

In this paper we propose a multivariate generalization of DerSimonian and Laird's procedure [11]. This is much simpler than REML and merely requires solving linear equations and standard matrix operations. In particular, in contrast to the alternatives, no

MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 0SR, U.K.
**Correspondence to: Dan Jackson, MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 0SR, U.K.*
†*E-mail: daniel.jackson@mrc-bsu.cam.ac.uk*

Contract/grant sponsor: UK Medical Research Council; contract/grant numbers: U.1052.00.006, U.1052.00.001

numerical maximization or iteration is needed. A further advantage of the proposed approach is that an estimate of the between-study variance matrix can be obtained without the assumption of normality. Although computationally intensive procedures are now more feasible than when DerSimonian and Laird originally proposed their method it remains the most popular approach and the only related multivariate generalization is due to Pinto *et al.* [15], who provide an estimate of the between-study variance matrix where the diagonal entries are the usual DerSimonian and Laird estimates.

Motivating examples for developing multivariate moments estimates are described in Section 6. A further motivation comes from an examination of the reviews by the Cochrane Collaboration, which frequently perform separate univariate meta-analyses on fairly large numbers of related outcomes using the same set of studies. A multivariate approach is desirable in such instances but requires the routine application of high dimensional meta-analyses. The existing standard iterative multivariate procedure, described in detail below, becomes computationally prohibitive in high dimensions.

The rest of the paper is set out as follows. In Section 2 the univariate random effects model is described and the standard approach of DerSimonian and Laird is summarized. In Section 3 the multivariate random effects model is described and in Section 4 the DerSimonian and Laird approach is extended to estimate the between-study variance matrix in the multivariate case. In Section 5 a simulation study is performed, where the multivariate DerSimonian and Laird procedure is compared to REML. In Section 6 the proposed methodology is applied to some example datasets and Section 7 explains how the methodology may be extended to include covariates. Section 8 summarizes our conclusions.

2. Univariate meta-analysis

The univariate random effects model assumes that the estimate of treatment effect from the ith study, y_i, is distributed as $Y_i|\mu_i \sim N(\mu_i, \sigma_i^2)$, where μ_i is the true underlying treatment effect of the ith study and σ_i^2 is the corresponding within-study variance. The variances σ_i^2 are unknown, but consistent estimates are used for these in practice, which are treated as known when pooling the studies' results. Estimates of within-study variance for a wide range of measures of treatment effect used in meta-analysis can be obtained as described by Sutton *et al.* [1]. The univariate random effects model further assumes that $\mu_i \sim N(\mu, \tau^2)$, where μ and τ^2 denote the overall treatment effect and between-study variance, respectively, and that the studies are independent. This provides the marginal model $Y_i \sim N(\mu, \sigma_i^2 + \tau^2)$. The main statistical difficulty lies in estimating τ^2.

The simplest and most commonly used estimate of τ^2 is due to DerSimonian and Laird [11]. This uses the heterogeneity statistic,

$$Q = \sum_{i=1}^{n} w_i (Y_i - \bar{Y})^2$$

where $w_i = \sigma_i^{-2}$, $\bar{Y} = \sum_{i=1}^{n} w_i Y_i / \sum_{i=1}^{n} w_i$ and n denotes the number of studies. Under the null hypothesis that $\tau^2 = 0$, Q follows a $\chi^2_{(n-1)}$ distribution and hence is a heterogeneity test statistic. Under the assumptions of the random effects model, it can be shown that the expectation of Q is

$$\mathrm{E}[Q] = (n-1) + \left(S_1 - \frac{S_2}{S_1} \right) \tau^2$$

where $S_r = \sum_{i=1}^{n} w_i^r$, which provides the DerSimonian and Laird moment estimator

$$\hat{\tau}^2 = \max\left(0, \frac{Q - (n-1)}{S_1 - \dfrac{S_2}{S_1}} \right)$$

Truncation at zero is to ensure that the variance estimate is non-negative. The estimate of the overall treatment effect is given by $\hat{\mu} = \sum_{i=1}^{n} w_i^* Y_i / \sum_{i=1}^{n} w_i^*$, where $w_i^* = (\sigma_i^2 + \hat{\tau}^2)^{-1}$, and the distribution of $\hat{\mu}$ is approximately $\hat{\mu} \sim N(\mu, (\sum_{i=1}^{n} w_i^*)^{-1})$. Confidence intervals and results from hypothesis tests are easily obtained, although it should be noted that this procedure only provides a good approximation when there is a sufficiently large number of relatively big studies [9].

3. Multivariate meta-analysis

We present the methods for a bivariate meta-analysis but the extension to the multivariate case is straightforward.

The bivariate random effects meta-analysis model assumes that study i provides two estimated treatment effects, x_i and y_i. These estimates may be correlated and it is assumed that

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \Bigg| \begin{pmatrix} \mu_{X_i} \\ \mu_{Y_i} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{X_i} \\ \mu_{Y_i} \end{pmatrix}, \begin{pmatrix} \sigma^2_{X_i} & \rho_i \sigma_{X_i} \sigma_{Y_i} \\ \rho_i \sigma_{X_i} \sigma_{Y_i} & \sigma^2_{Y_i} \end{pmatrix} \right) \tag{1}$$

where μ_{X_i} and μ_{Y_i} denote the true treatment effects in the ith study. The within-study variance matrices in (1) are estimated in practice but assumed fixed and known in application. Estimation of the within-study correlations ρ_i may require some innovation; it is easiest with individual patient data, because it may need information that is not provided in the published reports of individual studies. Formulae for estimating within-study covariances for some important special cases have been provided [16]. An alternative formulation of this type of model that does not require the within-study correlations to be estimated has also been proposed [17]. Although this is useful, it has been noted that a variety of types of errors in the estimation of these correlations generally have little impact on inferences concerning the treatment effect, if there is complete data (full set of estimated treatment effects and within-study variance matrices) [18].

The bivariate random effects model further assumes

$$\begin{pmatrix}\mu_{X_i}\\ \mu_{Y_i}\end{pmatrix} \sim N\left(\begin{pmatrix}\mu_X\\ \mu_Y\end{pmatrix}, \begin{pmatrix}\tau_X^2 & \kappa\tau_X\tau_Y\\ \kappa\tau_X\tau_Y & \tau_Y^2\end{pmatrix}\right) \tag{2}$$

where μ_X and μ_Y are the two overall treatment effects and are the parameters of central interest; the remaining three parameters in (2), τ_X^2, τ_Y^2 and κ, describe the between-study variation. Marginally, this provides the model

$$\begin{pmatrix}X_i\\ Y_i\end{pmatrix} \sim N\left(\begin{pmatrix}\mu_X\\ \mu_Y\end{pmatrix}, \begin{pmatrix}\sigma_{X_i}^2+\tau_X^2 & \rho_i\sigma_{X_i}\sigma_{Y_i}+\kappa\tau_X\tau_Y\\ \rho_i\sigma_{X_i}\sigma_{Y_i}+\kappa\tau_X\tau_Y & \sigma_{Y_i}^2+\tau_Y^2\end{pmatrix}\right) \tag{3}$$

The studies are assumed independent, and hence the likelihood can be obtained and maximized to provide maximum likelihood estimates of all parameters. It is however more usual to estimate the entries of the between-study variance matrix using REML. This estimation can easily be performed by maximizing a special case of the expression λ_{REML} given by Jennrich and Schluchter [19], page 812 (subject to the constraints that the between-study variance matrix is positive semi-definite). Specifically the expression to maximize is

$$\lambda_{\text{REML}} = -\frac{1}{2}\sum_{i=1}^{n}\log|\Sigma+\Delta_i| - \frac{1}{2}\log\left|\sum_{i=1}^{n}(\Sigma+\Delta_i)^{-1}\right| - \frac{1}{2}\sum_{i=1}^{n} r_i^{\mathrm{T}}(\Sigma+\Delta_i)^{-1}r_i \tag{4}$$

where Δ_i is the within-study variance matrix of the ith study,

$$\Delta_i = \begin{pmatrix}\sigma_{X_i}^2 & \rho_i\sigma_{X_i}\sigma_{Y_i}\\ \rho_i\sigma_{X_i}\sigma_{Y_i} & \sigma_{Y_i}^2\end{pmatrix}$$

Σ denotes the between-study variance matrix

$$\Sigma = \begin{pmatrix}\tau_X^2 & \kappa\tau_X\tau_Y\\ \kappa\tau_X\tau_Y & \tau_Y^2\end{pmatrix}$$

and the r_i denote the residuals, $r_i = (x_i, y_i)^{\mathrm{T}} - (\hat{\mu}_X, \hat{\mu}_Y)^{\mathrm{T}}$, where the vector of estimates $(\hat{\mu}_X, \hat{\mu}_Y)^{\mathrm{T}}$ is given by (5) below. The REML estimate of the between-study variance matrix will be denoted by Σ_{REML}.

3.1. Pooling the study estimates

Assuming all studies provide all outcomes (i.e. there are no missing data) then once the between-study variance matrix has been estimated, and irrespective of the estimation procedure used, the pooled estimates are given by [4]

$$\begin{pmatrix}\hat{\mu}_X\\ \hat{\mu}_Y\end{pmatrix} = \left(\sum_{i=1}^{n}(\hat{\Sigma}+\Delta_i)^{-1}\right)^{-1}\left(\sum_{i=1}^{n}(\hat{\Sigma}+\Delta_i)^{-1}\mathbf{Z}_i\right) \tag{5}$$

where $\mathbf{Z}_i$ denotes $(x_i, y_i)^{\mathrm{T}}$. The estimates (5) are approximately normally distributed with variance matrix

$$C = \mathrm{Var}(\hat{\mu}_X, \hat{\mu}_Y) = \left(\sum_{i=1}^{n}(\hat{\Sigma}+\Delta_i)^{-1}\right)^{-1} \tag{6}$$

An approximate $(1-\alpha)$ per cent confidence interval can be obtained for μ_X as $\hat{\mu}_X \pm Z_{\alpha/2}\sqrt{C_{(1,1)}}$, where $Z_{\alpha/2}$ denotes the $\alpha/2$ percentile of a normal distribution and $C_{(i,j)}$ denotes the entry in the ith row and jth column of C. The use of the t distribution with $n-1$ degrees of freedom, rather than the standard normal, has been suggested [5]. That is, intervals are obtained using $\hat{\mu}_X \pm t_{\alpha/2,n-1}\sqrt{C_{(1,1)}}$, where $t_{\alpha/2,v}$ denotes the $\alpha/2$ percentile of a t distribution with v degrees of freedom. This modified procedure will be considered in the simulation study below; in the univariate case neither null distribution consistently outperforms the other [9]. Alternative procedures using t distribution quantiles have also been developed [20–22].

3.2. Dealing with unreported outcomes

If some studies have missing outcomes then, assuming that these are missing at random, such studies can be incorporated into the matrix solutions (5) and (6) by allocating very large within-study variances to these missing observations; 10^{12} was used for this purpose in the simulation studies below, where the missing study outcomes and within-study correlations were set to 0. This replaces missing outcomes with estimates with negligible weight and information. These same values were used when estimating the between-study variance matrix using REML i.e. maximizing (4). An alternative however is to use the sum of (4) for studies providing all outcomes and the restricted log-likelihood of the marginal model for the others. The possibility of missing data is especially important, as previous articles highlight that there is little benefit of bivariate over univariate meta-analysis for complete data [5, 23].

4. Proposed new procedure for multivariate meta-analysis

The univariate procedure suggested by DerSimonian and Laird centres around the Q statistic. We propose, as a multivariate extension of this statistic, the matrix

$$Q=\begin{bmatrix} \sum_{i\in\mathbf{R}_X}\dfrac{(X_i-\bar{X}_1)^2}{\sigma^2_{X_i}} & \sum_{i\in\mathbf{R}_{X,Y}}\dfrac{(X_i-\bar{X}_2)(Y_i-\bar{Y}_2)}{\sigma_{X_i}\sigma_{Y_i}} \\ \sum_{i\in\mathbf{R}_{X,Y}}\dfrac{(X_i-\bar{X}_2)(Y_i-\bar{Y}_2)}{\sigma_{X_i}\sigma_{Y_i}} & \sum_{i\in\mathbf{R}_Y}\dfrac{(Y_i-\bar{Y}_1)^2}{\sigma^2_{Y_i}} \end{bmatrix}$$

where $\mathbf{R}_X$, $\mathbf{R}_Y$ and $\mathbf{R}_{X,Y}$ denote the sets of studies where X, Y and both outcomes X and Y are reported, respectively. The symbols $\bar{X}_1$ and $\bar{X}_2$ denote weighted averages of the X_i over the studies that report X, and those that report X *and* Y, where the weights are $\sigma^{-2}_{X_i}$ and $(\sigma_{X_i}\sigma_{Y_i})^{-1}$, respectively, that is

$$\bar{X}_1=\frac{\sum_{i\in\mathbf{R}_X}X_i/\sigma^2_{X_i}}{\sum_{i\in\mathbf{R}_X}1/\sigma^2_{X_i}}$$

and

$$\bar{X}_2=\frac{\sum_{i\in\mathbf{R}_{X,Y}}X_i/(\sigma_{X_i}\sigma_{Y_i})}{\sum_{i\in\mathbf{R}_{X,Y}}1/(\sigma_{X_i}\sigma_{Y_i})}$$

Similarly for $\bar{Y}_1$ and $\bar{Y}_2$,

$$\bar{Y}_1=\frac{\sum_{i\in\mathbf{R}_Y}Y_i/\sigma^2_{Y_i}}{\sum_{i\in\mathbf{R}_Y}1/\sigma^2_{Y_i}}$$

and

$$\bar{Y}_2=\frac{\sum_{i\in\mathbf{R}_{X,Y}}Y_i/(\sigma_{X_i}\sigma_{Y_i})}{\sum_{i\in\mathbf{R}_{X,Y}}1/(\sigma_{X_i}\sigma_{Y_i})}$$

Note that if all studies report both outcomes then all summations extend over all studies. When computing the matrix Q, missing outcomes can also be handled by replacing the missing data by arbitrary values with very large within-study variances and summing over all studies.

The diagonal entries of Q are the usual univariate 'Q' or 'χ^2 heterogeneity statistics' used by DerSimonian and Laird to estimate the between-study variance. Let n_X, n_Y and $n_{X,Y}$ denote the number of studies that provide X, Y and both outcomes, respectively. Further denote the expectation of Q by E. The standard univariate result gives

$$e_{(1,1)}=E[Q_{(1,1)}]=(n_X-1)+\left(\sum_{i\in\mathbf{R}_X}\sigma^{-2}_{X_i}-\sum_{i\in\mathbf{R}_X}\sigma^{-4}_{X_i}\Big/\sum_{i\in\mathbf{R}_X}\sigma^{-2}_{X_i}\right)\tau^2_X \tag{7}$$

and

$$e_{(2,2)}=E[Q_{(2,2)}]=(n_Y-1)+\left(\sum_{i\in\mathbf{R}_Y}\sigma^{-2}_{Y_i}-\sum_{i\in\mathbf{R}_Y}\sigma^{-4}_{Y_i}\Big/\sum_{i\in\mathbf{R}_Y}\sigma^{-2}_{Y_i}\right)\tau^2_Y \tag{8}$$

These expectations are both linear functions of just one of the unknown between-study variances. Hence we may match $q_{(1,1)}=e_{(1,1)}$ and $q_{(2,2)}=e_{(2,2)}$ and solve the linear equations to estimate τ^2_X and τ^2_Y, exactly as in the univariate case.

We show in Appendix A that

$$e_{(2,1)} = e_{(1,2)} = a + b\kappa\tau_X\tau_Y \tag{9}$$

where

$$a = \sum_{i\in\mathbf{R}_{X,Y}} \rho_i - \frac{\sum_{i\in\mathbf{R}_{X,Y}} \frac{\rho_i}{\sigma_{X_i}\sigma_{Y_i}}}{\sum_{i\in\mathbf{R}_{X,Y}} \frac{1}{\sigma_{X_i}\sigma_{Y_i}}} \tag{10}$$

and

$$b = \sum_{i\in\mathbf{R}_{X,Y}} \frac{1}{\sigma_{X_i}\sigma_{Y_i}} - \frac{\sum_{i\in\mathbf{R}_{X,Y}} \frac{1}{\sigma^2_{X_i}\sigma^2_{Y_i}}}{\sum_{i\in\mathbf{R}_{X,Y}} \frac{1}{\sigma_{X_i}\sigma_{Y_i}}} \tag{11}$$

Since the expectation $e_{(2,1)} = e_{(1,2)}$ is a linear function in the between-study covariance $\kappa\tau_X\tau_Y$, this covariance can also be estimated by equating $q_{(2,1)} = e_{(2,1)}$ in order to estimate the entire term $\kappa\tau_X\tau_Y$. Hence estimates of all entries of the between-study variance matrix can be obtained by matching moments, providing a symmetrical estimated 'untruncated' between-study variance matrix which we call Σ_{DL}.

4.1. Additional dimensions

Additional dimensions are easily added to the bivariate model described here. For example, if each study provided three estimates of treatment effect, then the between-study variances τ^2_X and τ^2_Y are estimated by matching moments using (7) and (8), and the additional between-study variance is similarly estimated using an analogous additional moments argument. A similar comment applies to the estimates of the between-study covariances.

4.2. Finding a positive semi-definite between-study variance estimate

There is however the difficulty, just as in the univariate case, that between-study variance estimates need not lie in the parameter space, that is Σ_{DL} need not be positive semi-definite. To addresses this we write Σ_{DL} in terms of its spectral decomposition

$$\Sigma_{DL} = \sum_{i=1}^{k} \lambda_i e_i e_i^{\mathsf{T}} \tag{12}$$

where λ_i is the ith eigenvalue of Σ_{DL} and e_i is the corresponding normalized eigenvector; k denotes the dimension of the multivariate meta-analysis, so that, for example, in the bivariate case $k=2$. We suggest using

$$\Sigma_{DL^+} = \sum_{i=1}^{k} \max(0, \lambda_i) e_i e_i^{\mathsf{T}} \tag{13}$$

to produce a 'truncated' symmetric and positive semi-definite estimate of Σ. This is a matrix with the same eigenvectors as the 'untruncated' version (12) but where any negative eigenvalues of (12) have been replaced by zero. In the univariate case, this procedure reduces to the truncation of the variance estimate at zero.

Estimates of overall treatment effect and its variance matrix are given by (5) and (6) with Σ replaced by Σ_{DL^+}, and the variance matrix of this estimate is similarly obtained from (6) upon making this same substitution. Confidence intervals may be constructed as described above. A random effects model collapses to a fixed effects analysis when all eigenvalues of Σ_{DL} are non-positive which in turn implies that all diagonal entries of Σ_{DL} are also non-positive, as all diagonal entries of a real symmetric square matrix are less than or equal to the largest eigenvalue [24]. Hence collapsing to a fixed effects analysis implies that the diagonal entries of Q, the usual univariate heterogeneity test statistics, are less than or equal to their associated degrees of freedom.

5. Simulation study

5.1. Study design

We compare the proposed multivariate extension of the DerSimonian and Laird procedure with the more conventional REML procedure. Bivariate meta-analyses with a sample size of $n=10$ were initially considered, which provides a moderate but not unreasonable number of studies. The parameters ρ_i, τ^2_X, τ^2_Y and κ were varied in the simulation study to cover a wide range of plausible circumstances.

For each simulation, two sets of 10 within-study variances were simulated from $0.25 \times \chi^2_1$, but where values outside the range [0.009, 0.6] were discarded, a procedure that follows Brockwell and Gordon [25]. These two sets of within-study variances were then ranked, and the first study was taken to have the largest pair of simulated values as $\sigma^2_{X_i}$ and $\sigma^2_{Y_i}$, and so on until the last study had the smallest pair of simulated within-study variances. This procedure simulates a mixture of study sizes. New within-study variances were simulated for every meta-analysis in the simulation study.

Pairs of x_i and y_i were then simulated directly from (3); the true overall treatment effects μ_X and μ_Y were both taken to be zero, although this choice is immaterial as all values tabulated below, with the exception of point estimates which are simply translated when adopting alternative true effects, are location invariant. The between-study variance matrix for each simulated dataset was then estimated using both REML and the proposed multivariate version of the DerSimonian and Laird procedure; the corresponding estimates of the treatment effects μ_X and μ_Y and their variances were obtained from (5) and (6) by replacing Σ by the appropriate estimate. Confidence intervals for the treatment effects were calculated using the two methods as described in Section 3.1. All the necessary simulations and calculations were performed using *R* version 2.5.1. The convergence of the REML procedure was checked using the diagnostic *convergence* of the *optim* command; datasets were repeatedly given new starting values until they passed *optim*'s default convergence diagnostics. The maximization of the restricted likelihood was performed using the logarithm of the between-study variances and Fisher's transformation of the between-study correlation.

We summarized the simulation study using the estimates' means and the proportions of confidence intervals that include the true treatment effect. Riley *et al.* [5] use 1000 simulations in their simulation studies, and this same number is used here because the REML procedure is computationally expensive and all that is desired is a comparison between the two competing procedures. Both methods are applied to the same set of 25 000 simulated meta-analyses and this provides a large number of estimates to compare. If a very accurate assessment of the coverage probability of confidence intervals was required then further simulations could be used to reduce the Monte Carlo error.

In order to choose suitable parameter values to investigate in the simulation study, Higgins and Thompson's [26] typical within-study variance, σ^2_t, was considered. This is given by

$$\sigma^2_t = \frac{(n-1)\sum_{i=1}^n w_i}{(\sum_{i=1}^n w_i)^2 - \sum_{i=1}^n w_i^2}$$

where $w_i = \sigma_i^{-2}$. With the σ^2_i generated as described above, we calculate $\sigma^2_t \to 0.056$ as $n \to \infty$. Then $I^2_X = \tau^2_X/(0.056+\tau^2_X)$ and $I^2_Y = \tau^2_Y/(0.056+\tau^2_Y)$ give the proportion of marginal variation in X and Y due to heterogeneity [26]. Three values of these I^2 terms were considered: 0 (no marginal between-study heterogeneity), 0.3 (mild heterogeneity) and 0.75 (notable heterogeneity), giving nine pairs of I^2 values. In simulation runs 1–9 these combinations of values were considered with both the between-study correlation κ and all within-study correlations ρ_i set to 0, in order to investigate the special case where all outcomes are independent. Where one or both of the I^2 values is 0 the correlation κ is not defined, and $\kappa = 0$ is taken to mean that the covariance $\kappa\tau_X\tau_Y = 0$ in such instances.

Runs 10–17 considered situations where the between and within-study correlations are similar, where κ and all ρ_i were set to 0.7 or to 0.95; only the combinations of $I^2_X = (0.3, 0.75)$ and $I^2_Y = (0.3, 0.75)$ were considered when using these correlations, as I^2 values of zero do not permit such a correlation.

The sensitivity and specificity of diagnostic tests provide an important scenario where within-study correlations are known to be zero, but where between-study correlation is anticipated [5]. Runs 18–25 therefore repeated runs 10–17 with the ρ_i all set to zero.

5.2. Results

The principal results from this simulation study are shown in Table I; the same data were used for both methods resulting in small paired differences within simulation runs. The table shows inferences relating to the first treatment effect μ_X; inferences for μ_Y can be obtained by symmetry. Columns 1–5 of Table I describe the parameter values used and the next two columns provide the mean estimates of μ_X, where 'DL' and 'RE' refer to the multivariate DerSimonian and Laird and REML procedures, respectively. The next two columns give the Monte Carlo (MC) variances of these estimates showing that the methods have equal precision. There was no evidence of bias (MC standard errors of $E(\hat{\mu}_X)$ were between 0.003 and 0.006, results not shown). The mean length of nominal 95 per cent confidence intervals using the standard normal quantile in the usual way, that is the mean length of the interval $\hat{\mu}_X \pm Z_{0.025}\sqrt{C_{(1,1)}}$, are also given in Table I. The length of the corresponding intervals resulting from using the t distribution can be obtained as the product of this tabulated value and $t_{0.025,9}/Z_{0.025} \approx 1.154$. Finally, Table I gives the estimated actual coverage probabilities of nominal 95 per cent confidence intervals using the normal and t distributions.

Table I shows that the multivariate DerSimonian and Laird procedure performs similarly to REML. When there is no heterogeneity, both procedures provide actual coverage probabilities of more than the nominal 95 per cent as expected, even when using the standard normal quantile to construct intervals, as estimates of between-study variance can only be overestimated. When the heterogeneity becomes notable however the t distribution becomes more suitable, again as expected [9, 27]. That the multivariate DerSimonian and Laird procedure is less computationally demanding is emphasized by the computing times required: using the DerSimonian and Laird approach all simulated datasets and fitted models required to produce Table I were produced in a few minutes; for REML the entire procedure was an overnight task.

Table I. Properties of estimates $\hat{\mu}_X$ from the simulation study with $n=10$, using 1000 simulations for each run.

					$E(\hat{\mu}_X)$		$\text{Var}(\hat{\mu}_X)$		Z length		Z coverage		t coverage	
Run	I^2_X	I^2_Y	κ	ρ_i	DL	RE	DL	RE	DL	RE	DL	RE	DL	RE
1	0	0	0	0	0.002	0.001	0.007	0.007	0.370	0.355	0.961	0.960	0.987	0.977
2	0	0.3	0	0	−0.001	−0.001	0.008	0.008	0.372	0.357	0.960	0.952	0.980	0.977
3	0	0.75	0	0	0.000	0.000	0.008	0.008	0.366	0.356	0.962	0.965	0.988	0.983
4	0.3	0	0	0	−0.005	−0.004	0.012	0.011	0.418	0.408	0.936	0.928	0.966	0.960
5	0.3	0.3	0	0	0.007	0.007	0.011	0.011	0.413	0.401	0.941	0.925	0.963	0.950
6	0.3	0.75	0	0	−0.003	−0.002	0.012	0.012	0.420	0.414	0.929	0.919	0.957	0.947
7	0.75	0	0	0	0.006	0.006	0.029	0.028	0.630	0.631	0.912	0.913	0.950	0.945
8	0.75	0.3	0	0	−0.010	−0.010	0.029	0.029	0.631	0.630	0.895	0.892	0.930	0.924
9	0.75	0.75	0	0	−0.004	−0.004	0.028	0.028	0.631	0.628	0.916	0.915	0.942	0.936
10	0.3	0.3	0.7	0.7	0.006	0.006	0.011	0.011	0.406	0.408	0.927	0.925	0.958	0.959
11	0.3	0.75	0.7	0.7	−0.004	−0.004	0.012	0.012	0.403	0.402	0.919	0.925	0.953	0.957
12	0.75	0.3	0.7	0.7	0.005	0.005	0.030	0.030	0.636	0.642	0.908	0.914	0.942	0.945
13	0.75	0.75	0.7	0.7	−0.005	−0.004	0.030	0.030	0.626	0.629	0.892	0.885	0.932	0.934
14	0.3	0.3	0.95	0.95	−0.001	−0.001	0.010	0.010	0.384	0.381	0.909	0.911	0.948	0.953
15	0.3	0.75	0.95	0.95	0.002	0.002	0.011	0.010	0.392	0.392	0.938	0.949	0.963	0.970
16	0.75	0.3	0.95	0.95	0.001	0.001	0.030	0.030	0.603	0.623	0.889	0.905	0.918	0.933
17	0.75	0.75	0.95	0.95	−0.001	−0.001	0.028	0.028	0.613	0.619	0.890	0.893	0.931	0.929
18	0.3	0.3	0.7	0	0.003	0.003	0.012	0.012	0.416	0.410	0.922	0.915	0.956	0.948
19	0.3	0.75	0.7	0	0.001	0.001	0.011	0.011	0.413	0.409	0.932	0.922	0.961	0.959
20	0.75	0.3	0.7	0	0.008	0.008	0.029	0.029	0.645	0.647	0.908	0.910	0.938	0.939
21	0.75	0.75	0.7	0	−0.001	−0.001	0.028	0.028	0.616	0.622	0.903	0.906	0.938	0.935
22	0.3	0.3	0.95	0	0.003	0.003	0.011	0.011	0.415	0.410	0.930	0.927	0.963	0.956
23	0.3	0.75	0.95	0	−0.002	−0.002	0.011	0.011	0.416	0.410	0.939	0.935	0.958	0.962
24	0.75	0.3	0.95	0	0.004	0.004	0.031	0.031	0.638	0.640	0.901	0.902	0.935	0.938
25	0.75	0.75	0.95	0	0.005	0.006	0.029	0.029	0.639	0.640	0.909	0.920	0.952	0.958

In each case, DL and RE denote values using the multivariate DerSimonian and Laird and REML procedures, respectively. $E(\hat{\mu}_X)$ denotes the average estimated first treatment effect and $\text{Var}(\hat{\mu}_X)$ denotes the Monte Carlo variance of these estimates. 'Z length' is the average length of a nominal 95 per cent confidence interval for the first treatment effect using the standard normal quantile and 'Z coverage' and 't coverage' denote the proportion of nominal 95 per cent confidence intervals that cover the true value, using the standard normal and t quantiles, respectively.

Table II. Properties of estimates of Σ from the simulation study with $n=10$, using 1000 simulations for each run.

					$E(\hat{\tau}_X)$		$E(\hat{\text{Cov}})$		$e=2$		$e=1$		$e=0$		$\kappa\approx\pm1$	
Run	τ_X^2	τ_Y^2	$\kappa\tau_X\tau_Y$	ρ_i	DL	RE	DL	RE	DL	RE	DL	RE	DL	RE	DL	RE
1	0	0	0	0	0.018	0.015	0.000	0.003	91	28	671	529	238	443	520	219
2	0	0.024	0	0	0.020	0.016	0.000	0.004	169	76	705	637	126	287	564	317
3	0	0.168	0	0	0.018	0.017	0.002	0.005	286	177	692	760	22	63	568	433
4	0.024	0	0	0	0.038	0.037	0.000	0.004	159	75	733	643	108	282	582	320
5	0.024	0.024	0	0	0.035	0.033	0.000	0.003	239	146	693	665	68	189	589	377
6	0.024	0.168	0	0	0.037	0.037	0.002	0.005	464	367	529	604	7	29	462	430
7	0.168	0	0	0	0.166	0.167	0.000	0.005	292	183	692	761	16	56	563	445
8	0.168	0.024	0	0	0.167	0.167	0.000	0.001	433	342	561	626	6	32	476	438
9	0.168	0.168	0	0	0.168	0.167	−0.001	0.000	781	682	218	316	1	2	195	260
10	0.024	0.024	0.017	0.7	0.035	0.037	0.021	0.025	242	163	683	714	75	123	566	521
11	0.024	0.168	0.045	0.7	0.035	0.036	0.044	0.052	415	335	581	648	4	17	500	501
12	0.168	0.024	0.045	0.7	0.176	0.179	0.048	0.054	454	346	539	640	7	14	471	482
13	0.168	0.168	0.118	0.7	0.170	0.172	0.117	0.119	762	702	238	294	0	4	230	250
14	0.024	0.024	0.023	0.95	0.035	0.035	0.030	0.032	222	164	742	755	36	81	613	625
15	0.024	0.168	0.060	0.95	0.035	0.037	0.061	0.070	376	294	624	705	0	1	575	624
16	0.168	0.024	0.060	0.95	0.171	0.183	0.062	0.075	348	301	650	695	2	4	597	616
17	0.168	0.168	0.160	0.95	0.171	0.173	0.160	0.164	623	583	376	416	1	1	405	387
18	0.024	0.024	0.017	0	0.038	0.037	0.015	0.018	237	153	699	681	64	166	593	490
19	0.024	0.168	0.045	0	0.039	0.038	0.042	0.043	364	299	625	674	11	27	573	546
20	0.168	0.024	0.045	0	0.179	0.182	0.044	0.045	374	282	615	692	11	26	562	563
21	0.168	0.168	0.118	0	0.160	0.166	0.105	0.109	606	527	391	466	3	7	384	427
22	0.024	0.024	0.023	0	0.036	0.036	0.018	0.021	191	98	735	723	74	179	639	550
23	0.024	0.168	0.060	0	0.041	0.040	0.056	0.058	288	214	692	741	20	45	650	649
24	0.168	0.024	0.060	0	0.174	0.175	0.054	0.056	301	222	693	744	6	34	645	646
25	0.168	0.168	0.160	0	0.184	0.184	0.161	0.161	371	288	628	705	1	7	627	691

In each case, DL and RE denote values using the multivariate DerSimonian and Laird and REML procedures, respectively. $E(\hat{\tau}_X^2)$ denotes the average estimated between-study variance and $E(\hat{\text{Cov}})$ denotes the average estimate of the covariance $\kappa\tau_X\tau_Y$. $e=2,1,0$ refer to the number of simulated meta-analyses that fit random effects models with variance matrices with 2 (full random effects), 1 (degenerate random effect) and 0 (fixed effects) positive eigenvalues, respectively. $\kappa\approx\pm1$ is the number of simulated meta-analyses that effectively provide an estimate of κ that lies at the edge of its parameter space.

The average estimates of τ_X^2 and covariance $\kappa\tau_X\tau_Y$ are shown for all 25 runs in Table II. Although there is little to choose between the two methods, the results suggest that the multivariate DerSimonian and Laird procedure slightly further overestimates between-study variances when the data are in fact homogeneous.

Table II also shows the number of simulations with a estimated between-study variance matrix with two, one or no positive eigenvalues ($e=2,1,0$, respectively), meaning that a full bivariate random effects model, a degenerate random effects model or a fixed effects model has effectively been fitted, respectively. The number of positive eigenvalues for the multivariate DerSimonian and Laird procedure is immediate from (13). When computing these values for REML, estimated between-study variances below 0.0005 were regarded as 0, and estimated between-study correlations with magnitude above 0.995 were regarded as 1. A between-study variance of 0.0005 corresponds to $I^2 \approx 0.01$; stricter rounding, at 0.00005 and 0.9995, only changed the results very slightly.

There are two ways in which a between-study variance matrix could provide $e=1$: either the between-study correlation could have magnitude one, or a between-study variance is zero. In order to investigate the reasons for a degenerate random effect, the final column of Table II gives the number of simulated datasets that effectively estimate $\hat{\kappa}=\pm 1$, i.e. provide between-study variances that are both greater than 0.0005 and $|\hat{\kappa}|>0.995$. Note that these requirements, imposed due to the numerical maximization required when performing REML, means that, when using the DerSimonian and Laird method, more simulations could meet these requirements than provide $e=1$. This only happened in run 17 which provides a rather extreme set of circumstances however. Without this rounding, all truncated DerSimonian and Laird estimates with $e=1$ imply $\hat{\kappa}=1$ because positive estimated between-study variances are inevitable when $e=1$ due to the nature of the truncation used. Table II confirms the previous finding that estimates of between-study correlation frequently lie at the edge of the parameter space [5]; however the two competing methods perform fairly similarly in almost every respect. The large number of simulated meta-analyses with κ at the edge of its parameter space explains the positive bias in estimates of between-study variance parameters [5], which results in conservative standard errors. Despite this, the low coverage rates obtained when using the standard normal quantile, when the data is heterogeneous, carry over from the univariate case.

Sidik and Jonkman [28] also performed a simulation study comparing various estimators of the between-study variance in the univariate case. They concluded that heterogeneity estimates, including DerSimonian and Laird and REML, *underestimate* the heterogeneity variance, with *increasing bias as this variance increases*, and refer to a number of other simulation studies that agree with this. Unlike the simulation study performed here, and that of Riley *et al.* [5], these studies do not simulate under the assumptions of the random effects model; for example Sidik and Jonkman [28] simulate two by two tables and base all calculations on the resulting empirical log odds ratios and their estimated within-study variances. The simulation studies of Sidik and Jonkman and others suggest that if bivariate simulations were performed that reflect the approximate nature of the random effects model, conclusions concerning the estimation of the between-study variance matrix could be rather different.

5.3. Further simulation studies

Three further simulation studies were performed. First of all, like Riley *et al.* [5] we considered the scenario where half the studies have one of the outcomes missing completely at random. This and the original scenario were then repeated using $n=50$.

There was no evidence of bias in estimates of treatment effect and the proposed multivariate DerSimonian and Laird and REML procedures performed very similarly in every scenario and in every respect. For example, some indicative results from an alterative scenario are shown in Table III, where we consider the case where $n=10$ and half of the studies' first estimate of treatment effect are missing in every meta-analysis. Here the first two columns of results show the coverage probabilities of nominal 95 per cent confidence intervals for the first treatment effect (for which half the estimates are missing) using standard normal quantiles, and the next two columns show the corresponding results for the second treatment effect. The same results using a t distribution are also shown and finally the last two columns show the proportion of meta-analyses that effectively provide an estimate of κ that lies at the edge of its parameter space. Comparing these results to those shown in Table I, we see that the coverage probabilities of nominal 95 per cent confidence intervals for the first treatment effect, when the data is very heterogeneous, and using the standard normal quantile, drop further in a similar manner when using both methods and data are missing in this way. Furthermore, from a comparison with Table II, we see that the proportion of meta-analyses that provide $\hat{\kappa}\approx\pm 1$ increases now that the number of studies that provide information concerning this parameter is reduced, but both methods are affected similarly by the loss of information.

Table IV shows some of the benefits of the larger sample size of $n=50$ and complete data. The first six columns of results show the mean length of nominal 95 per cent confidence intervals using the standard normal quantile and the coverage probability of confidence intervals. A comparison with Table I shows that these intervals are much shorter due to the extra information and that the coverage probabilities are closer to the nominal level due to the larger sample size. The last two columns of Table IV show the number of meta-analyses than provide $e=0$. A comparison with Table II shows that, for example, the model is much less likely to collapse to a fixed effects analysis when there is heterogeneity with this larger sample size. Both methods benefit similarly from this extra information however. Results for all the additional simulation studies are available from the first author upon request.

Finally, a single very high dimensional dataset was generated, in order to illustrate the computational advantages of the proposed approach. One hundred studies, with complete data on fifteen outcomes, were simulated assuming all within-study correlations are zero and that all between-study variances and correlations are 0.024 and 0.7, respectively. Both approaches provided similar results but REML, implemented using the new *Stata* command *mvmeta* [29], took 12 hours to converge on a UNIX system; by comparison the proposed approach took around 2 seconds on a laptop.

Table III. Properties of confidence intervals from the simulation study with $n=10$, *but where half of the first outcomes are missing completely at random*, using 1000 simulations for each run.

	Z coverage (X)		Z coverage (Y)		t coverage (X)		t coverage (Y)		$\kappa \approx \pm 1$	
Run	DL	RE	DL	RE	DL	RE	DL	RE	DL	RE
1	0.960	0.956	0.977	0.967	0.997	0.996	0.989	0.983	606	257
2	0.968	0.962	0.933	0.912	0.992	0.990	0.953	0.935	643	375
3	0.952	0.943	0.897	0.884	0.995	0.994	0.934	0.926	661	587
4	0.932	0.922	0.969	0.960	0.986	0.984	0.985	0.984	637	333
5	0.922	0.920	0.942	0.924	0.987	0.984	0.970	0.956	635	422
6	0.923	0.907	0.916	0.903	0.981	0.974	0.946	0.930	612	555
7	0.870	0.867	0.967	0.951	0.948	0.947	0.986	0.977	703	491
8	0.869	0.856	0.930	0.916	0.946	0.943	0.964	0.957	638	492
9	0.850	0.842	0.907	0.905	0.952	0.936	0.937	0.924	474	492
10	0.951	0.948	0.939	0.919	0.989	0.984	0.968	0.952	715	548
11	0.937	0.935	0.925	0.911	0.983	0.982	0.946	0.939	690	649
12	0.872	0.872	0.943	0.925	0.946	0.949	0.975	0.957	680	574
13	0.885	0.871	0.910	0.905	0.952	0.946	0.949	0.930	527	549
14	0.938	0.916	0.936	0.918	0.988	0.977	0.964	0.954	787	679
15	0.916	0.937	0.910	0.913	0.987	0.988	0.940	0.937	759	677
16	0.891	0.901	0.950	0.935	0.956	0.962	0.968	0.969	724	666
17	0.864	0.878	0.898	0.897	0.943	0.960	0.935	0.925	646	606
18	0.940	0.932	0.956	0.928	0.985	0.981	0.970	0.958	660	467
19	0.925	0.926	0.912	0.906	0.985	0.983	0.951	0.945	657	624
20	0.844	0.845	0.939	0.927	0.940	0.931	0.963	0.953	663	603
21	0.878	0.880	0.914	0.907	0.953	0.951	0.949	0.947	539	610
22	0.933	0.931	0.933	0.906	0.987	0.982	0.963	0.943	659	509
23	0.943	0.935	0.921	0.909	0.989	0.990	0.949	0.941	681	630
24	0.872	0.869	0.941	0.917	0.960	0.958	0.967	0.950	705	616
25	0.879	0.881	0.922	0.911	0.955	0.953	0.948	0.937	638	724

DL and RE denote values using the multivariate DerSimonian and Laird and REML procedures, respectively. 'Z coverage (X)' and 'Z coverage (Y)' denote the proportion of nominal 95 per cent confidence intervals that cover the first and second true treatment effect. 't coverage' denotes these proportions using t quantiles and $\kappa \approx \pm 1$ is the number of simulated meta-analyses that effectively provide an estimate of κ that lies at the edge of its parameter space.

6. Examples

6.1. Sensitivity and specificity of tumour markers

Riley *et al.* [5] describe a systematic review concerning the sensitivity and specificity of tumour markers used for diagnosing primary bladder cancer. One of these markers was telomerase, a ribonucleoprotein enzyme, evaluated in 10 studies as shown in their Table I. The data comprised logit sensitivity and specificity, and their standard errors. The within-study correlations are zero in this situation. Estimating the between-study variance matrix using REML provides

$$\Sigma_{\text{REML}} = \begin{pmatrix} 0.202 & -0.723 \\ -0.723 & 2.584 \end{pmatrix}$$

with estimated overall means of 1.166 and 2.058, with standard errors of 0.186 and 0.561, respectively [29].

Using the proposed multivariate DerSimonian and Laird procedure gives

$$\Sigma_{DL^+} = \begin{pmatrix} 0.200 & -0.668 \\ -0.668 & 2.233 \end{pmatrix}$$

with estimated means of 1.166 and 2.030, and with standard errors of 0.186 and 0.520, respectively. These inferences agree to 3 decimal places for the first outcome (sensitivity) and are in good agreement for the second (specificity). A plot of the data, showing each of the 10 estimates and 95 per cent confidence ellipses, assuming within-study normal distributions, is shown in Figure 1. The solid point and curve are the estimated overall mean and the corresponding 95 per cent confidence ellipse using

Table IV. Properties of confidence intervals for the first treatment effect and fitted models from the simulation study with $n=50$.

	Z length		Z coverage		t coverage		$e=0$	
Run	DL	RE	DL	RE	DL	RE	DL	RE
1	0.147	0.140	0.959	0.952	0.961	0.958	164	486
2	0.145	0.138	0.970	0.957	0.973	0.960	27	113
3	0.144	0.138	0.957	0.955	0.960	0.958	0	0
4	0.177	0.175	0.943	0.933	0.952	0.940	22	89
5	0.176	0.169	0.943	0.932	0.950	0.937	5	30
6	0.174	0.172	0.940	0.933	0.950	0.939	0	0
7	0.295	0.295	0.945	0.948	0.954	0.956	0	0
8	0.294	0.294	0.950	0.948	0.955	0.956	0	0
9	0.294	0.295	0.951	0.949	0.956	0.952	0	0
10	0.174	0.172	0.931	0.931	0.941	0.936	3	10
11	0.173	0.172	0.939	0.938	0.945	0.945	0	0
12	0.288	0.288	0.936	0.940	0.944	0.945	0	0
13	0.294	0.294	0.936	0.941	0.939	0.948	0	0
14	0.171	0.169	0.944	0.940	0.949	0.947	1	4
15	0.165	0.166	0.928	0.938	0.935	0.940	0	0
16	0.275	0.276	0.936	0.942	0.939	0.948	0	0
17	0.291	0.293	0.923	0.927	0.934	0.932	0	0
18	0.178	0.176	0.939	0.932	0.943	0.944	3	19
19	0.173	0.171	0.939	0.937	0.941	0.947	0	0
20	0.292	0.293	0.955	0.959	0.960	0.961	0	0
21	0.290	0.290	0.949	0.949	0.951	0.953	0	0
22	0.176	0.175	0.936	0.930	0.943	0.936	2	29
23	0.173	0.171	0.943	0.945	0.946	0.948	0	0
24	0.291	0.292	0.948	0.946	0.951	0.949	0	0
25	0.286	0.285	0.946	0.947	0.955	0.955	0	0

DL and RE denote values using the multivariate DerSimonian and Laird and REML procedures, respectively. 'Z length' is the average length of a nominal 95 per cent confidence interval for the first treatment effect using the standard normal quantile and 'Z coverage' and 't coverage' denote the proportion of nominal 95 per cent confidence intervals that cover the true value, using the standard normal and t quantiles, respectively. $e=0$ is the number of simulated meta-analyses that collapse to a fixed effects model.

the multivariate DerSimonian and Laird procedure; a very similar, but not quite indistinguishable, confidence region is obtained using REML.

6.2. *Fibrinogen studies collaboration*

The Fibrinogen studies collaboration is a meta-analysis of individual data on 154 012 adults from 31 prospective studies with information on plasma fibrinogen and major disease outcomes [30]. As part of the published analysis, the incidence of coronary heart disease was compared across groups defined by baseline levels of fibrinogen [31]. That analysis used a fixed effects model; here we allow for heterogeneity between studies using a random-effects model and five groups, as presented previously by White [29]. This random effects model involves four correlated outcomes, which are the log hazard ratios, of the second to fifth groups relative to the first, adjusted for age and stratified by sex and study.

Estimating the between-study variance matrix using REML provides

$$\Sigma_{\text{REML}} = \begin{pmatrix} 0.052 & 0.064 & 0.068 & 0.053 \\ 0.064 & 0.082 & 0.088 & 0.075 \\ 0.068 & 0.088 & 0.095 & 0.086 \\ 0.053 & 0.075 & 0.086 & 0.107 \end{pmatrix}$$

with estimated log hazard ratios of 0.162, 0.393, 0.562 and 0.897, and with standard errors of 0.080, 0.088, 0.091 and 0.094, respectively [29].

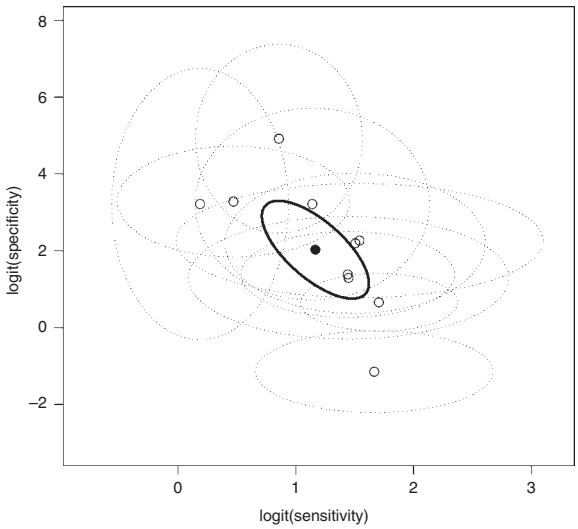

Figure 1. The sensitivity and specificity of tumour markers in 10 studies. The solid point and ellipsoid shows the estimate of the overall mean and the corresponding 95 per cent confidence ellipsoid using the multivariate DerSimonian and Laird procedure.

The multivariate DerSimonian and Laird procedure gives the somewhat different estimated between-study variance matrix of

$$\Sigma_{DL^+} = \begin{pmatrix} 0.030 & 0.043 & 0.050 & 0.038 \\ 0.043 & 0.063 & 0.073 & 0.068 \\ 0.050 & 0.073 & 0.085 & 0.077 \\ 0.038 & 0.068 & 0.077 & 0.126 \end{pmatrix}$$

with estimated log hazard ratios of 0.176, 0.405, 0.565 and 0.907, and with standard errors of 0.067, 0.077, 0.084 and 0.094, respectively. Although the resulting inferences are in reasonable agreement, slightly larger estimated log hazard ratios are provided by the DerSimonian and Laird procedure.

7. Meta-regression

In this section we extend univariate meta-regression, using the standard moments estimate of the between-study variance, to the multivariate case.

7.1. Univariate meta-regression

In univariate meta-regression we assume that the estimates of treatment effect are distributed as $Y_i \sim N(\mathbf{W}_i\beta, \sigma_i^2+\tau^2)$, where $\mathbf{W}_i$ denotes the row vector containing the covariate values associated with the ith study (i.e. $\mathbf{W}_i$ is the ith row of the regression 'design matrix' $\mathbf{W}$). The univariate Q meta-regression statistic is [22]

$$Q = \sum_{i=1}^{n_Y} \sigma_i^{-2}(Y_i - \mathbf{W}_i\hat{\beta}_{Y,1})^2 \tag{14}$$

where $\hat{\beta}_{Y,1}$ denotes the maximum likelihood estimate of β under the fixed effects assumption that $\tau^2=0$. The summation is over the n_Y studies that provide the outcome Y. The expectation of Q is given by [22]

$$E[Q] = (n_Y - r) + (\mathrm{tr}(\Lambda^{-1}) - \mathrm{tr}((\mathbf{W}^{\mathrm{T}}\Lambda^{-1}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\Lambda^{-2}\mathbf{W}))\tau^2$$

where r is the number of regression parameters, including the intercept, and $\Lambda = \mathrm{diag}(\sigma_i^2)$. This provides an estimate of τ^2 by equating Q to its expectation.

7.2. Multivariate meta-regression

Again for simplicity the bivariate model will be presented, which is given by

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{W}_{X_i}\beta_X \\ \mathbf{W}_{Y_i}\beta_Y \end{pmatrix}, \begin{pmatrix} \sigma_{X_i}^2+\tau_X^2 & \rho_i\sigma_{X_i}\sigma_{Y_i}+\kappa\tau_X\tau_Y \\ \rho_i\sigma_{X_i}\sigma_{Y_i}+\kappa\tau_X\tau_Y & \sigma_{Y_i}^2+\tau_Y^2 \end{pmatrix} \right) \tag{15}$$

Here $\mathbf{W}_{X_i}$ and $\mathbf{W}_{Y_i}$ refer to the corresponding ith rows of the relevant design matrices. It is not assumed that these are same for both outcomes; indeed this is unlikely in practice.

We suggest using the univariate Q statistics (14) for the diagonal entries of the Q matrix, so that the expectation of each of these is a linear function of a between-study variance as before. For example this gives

$$q_{(1,1)} = \sum_{i \in \mathbf{R}_X} \sigma_{X_i}^{-2} (X_i - \mathbf{W}_{X_i} \hat{\beta}_{X,1})^2$$

and equating these diagonal entries to their expected values gives the corresponding estimates of the entries of the between-study variance matrix. This reduces to evaluating the diagonal entries of Q defined in Section 4, and matching moments using (7) and (8), if there are no covariates.

For the off diagonal entries of Q we suggest

$$q_{(1,2)} = q_{(2,1)} = \sum_{i \in \mathbf{R}_{X,Y}} (\sigma_{X_i} \sigma_{Y_i})^{-1} (X_i - \mathbf{W}_{X_i} \hat{\beta}_{X,2})(Y_i - \mathbf{W}_{Y_i} \hat{\beta}_{Y,2}) \tag{16}$$

Writing $\Psi = \text{diag}(1/(\sigma_{X_i} \sigma_{Y_i}))$, the estimates $\hat{\beta}_{X,2}$ and $\hat{\beta}_{Y,2}$ in (16) are obtained from fitting the models $\mathbf{X} \sim N(\mathbf{W}_X \beta_X, \Psi^{-1})$ and $\mathbf{Y} \sim N(\mathbf{W}_Y \beta_Y, \Psi^{-1})$, respectively. This reduces to the usual meta-regression Q statistic (14) in the univariate case (if $X_i = Y_i$ and $\sigma_{X_i} = \sigma_{Y_i}$ for all i) and, if there are no covariates, to the diagonal entry of the Q matrix proposed for multivariate meta-analysis in Section 4. In particular, note that $\hat{\beta}_{X,2}$ and $\hat{\beta}_{Y,2}$ are analogous to $\bar{X}_2$ and $\bar{Y}_2$, respectively. Hence the proposed multivariate meta-regression procedure simplifies to the usual DerSimonian and Laird univariate procedure, its multivariate extension developed here and the usual univariate meta-regression procedure, as appropriate.

As shown in Appendix B, the expectation of (16) is a linear function of $\kappa \tau_1 \tau_2$ alone, specifically

$$E[Q_{(1,2)}] = e_{(1,2)} = a + b\kappa\tau_1\tau_2 \tag{17}$$

where

$$a = \text{tr}(\mathbf{AP}) \tag{18}$$

and

$$b = \text{tr}(\mathbf{A}\Psi) \tag{19}$$

where $\mathbf{A} = \mathbf{A}_X^{\mathrm{T}} \mathbf{A}_Y$, $\mathbf{A}_X = \mathbf{I} - \mathbf{V}_X (\mathbf{V}_X^{\mathrm{T}} \mathbf{V}_X)^{-1} \mathbf{V}_X^{\mathrm{T}}$, $\mathbf{A}_Y = \mathbf{I} - \mathbf{V}_Y (\mathbf{V}_Y^{\mathrm{T}} \mathbf{V}_Y)^{-1} \mathbf{V}_Y^{\mathrm{T}}$, $\mathbf{V}_X = \Psi^{1/2} \mathbf{W}_X$, $\mathbf{V}_Y = \Psi^{1/2} \mathbf{W}_Y$ and $\mathbf{P}$ denotes $\text{diag}(\rho_i)$.

The matrices $\mathbf{A}$, Ψ and $\mathbf{P}$ are all functions of the known within-study variance matrices, and hence so are a and b in (17); solving $q_{(2,1)} = E[Q_{(2,1)}]$ yields estimates of the between-study covariance. Hence all entries of an untruncated between-study covariance matrix can be obtained, which can be truncated as in (13), and the meta-regression proceeds as a weighted linear regression model, where all weights are regarded as known.

8. Conclusions

The simple and natural multivariate extension of the usual univariate DerSimonian and Laird procedure has been found to perform satisfactorily compared to the more established REML based procedure, both in the context of a simulation study and when analysing real datasets. This is despite the fact that it is much less computationally demanding than REML: it is non-iterative, does not require any convergence diagnostics and the experience of running simulation studies shows that it provides inferences much more quickly. These computational advantages become much greater as the number of dimensions increases. The proposed approach has been implemented in *Stata* using *mvmeta* [29] and *R* code is also available for this purpose from the first author on request. For those who prefer to estimate the between-study variance matrix using likelihood based methods, the proposed method could be used to give starting values for the necessary iterations. It seems reasonable to expect that REML and multivariate DerSimonian and Laird are likely to agree most closely where there is a small number of dimensions and moderate to large between-study variation.

Recently, DerSimonian and Kacker [32] proposed a general method of moments estimate of the between-study variance in the univariate case and further examine estimates that are obtained using two iterations. Other estimates of the between-study variance have also been suggested [28, 33, 34] and it is of interest to investigate how these might also be adapted and applied in the multivariate setting. We note that the proposed Q matrix is not based on matrix operations so is not invariant to reparametrizations. For example, replacing X_i and Y_i by $(X_i - Y_i)$ and $(X_i + Y_i)$ would not give the same solution when using our proposed procedure. A matrix version of Q might be $\sum_{i=1}^{n} \Delta_i^{-1/2} (\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})^{\mathrm{T}} \Delta_i^{-1/2}$, and such alternative definitions require further investigation, but it is not clear that this will correctly allow for missing data.

The proposed methodology is based entirely on matching moments and gives an estimate of the between-study variance matrix without the assumption of normality. This can be regarded as an advantage of our method. If the results from the multivariate DerSimonian and Laird and REML procedures differ greatly then this may cast doubt on the validity of the usual normal assumption. Although non-normal distributions may be used in meta-analysis [35–37], here the usual normal distributions

were assumed in the REML analysis for both the within and between-study distributions. In addition, recent developments in the context of univariate meta-analysis take into account the fact that the within-study variances are not truly known [38] and these ideas could be extended to the multivariate case. Finally, the procedure for ensuring the positive semi-definiteness of estimates of between-study variance matrices is not unique and others might also be considered. However we believe we have produced a useful and computationally straightforward method.

Appendix A: Deriving the expectation $e_{(1,2)}=e_{(2,1)}$ for meta-analysis

We evaluate the expectation

$$e_{(1,2)}=e_{(2,1)}=E(Q_{(1,2)})$$

assuming model (3) applies and that the studies are independent. In particular this model implies that $\mathrm{Cov}(X_i, Y_i)=\rho_i\sigma_{X_i}\sigma_{Y_i}+\kappa\tau_X\tau_Y$ and that $\mathrm{Cov}(X_i, Y_j)=0$ for $i\neq j$. We have that

$$\begin{aligned}q_{(1,2)}=\sum_{i\in\mathbf{R}_{X,Y}}\frac{(X_i-\bar{X}_2)(Y_i-\bar{Y}_2)}{\sigma_{X_i}\sigma_{Y_i}}&=\sum_{i\in\mathbf{R}_{X,Y}}\frac{(X_i-\mu_X+\mu_X-\bar{X}_2)(Y_i-\mu_Y+\mu_Y-\bar{Y}_2)}{\sigma_{X_i}\sigma_{Y_i}}\\&=\sum_{i\in\mathbf{R}_{X,Y}}\frac{(X_i-\mu_X)(Y_i-\mu_Y)}{\sigma_{X_i}\sigma_{Y_i}}+\sum_{i\in\mathbf{R}_{X,Y}}\frac{(X_i-\mu_X)(\mu_Y-\bar{Y}_2)}{\sigma_{X_i}\sigma_{Y_i}}\\&\quad+\sum_{i\in\mathbf{R}_{X,Y}}\frac{(Y_i-\mu_Y)(\mu_X-\bar{X}_2)}{\sigma_{X_i}\sigma_{Y_i}}+\sum_{i\in\mathbf{R}_{X,Y}}\frac{(\mu_X-\bar{X}_2)(\mu_Y-\bar{Y}_2)}{\sigma_{X_i}\sigma_{Y_i}}\end{aligned}$$

Noting that $\bar{X}_2$ and $\bar{Y}_2$ are weighted means, where the weights are $1/(\sigma_{X_i}\sigma_{Y_i})$, the last three terms are of the same magnitude, although signs differ, and the expression simplifies to

$$q_{(1,2)}=\sum_{i\in\mathbf{R}_{X,Y}}\frac{(X_i-\mu_X)(Y_i-\mu_Y)}{\sigma_{X_i}\sigma_{Y_i}}-\sum_{i\in\mathbf{R}_{X,Y}}\frac{(\bar{X}_2-\mu_X)(\bar{Y}_2-\mu_Y)}{\sigma_{X_i}\sigma_{Y_i}}$$

The expectation $e_{(1,2)}=e_{(2,1)}$ therefore equals the difference between the expectations of these two summations. The expectation of the first summation is

$$\begin{aligned}E\left(\sum_{i\in\mathbf{R}_{X,Y}}\frac{(X_i-\mu_X)(Y_i-\mu_Y)}{\sigma_{X_i}\sigma_{Y_i}}\right)&=\sum_{i\in\mathbf{R}_{X,Y}}\frac{\mathrm{Cov}(X_i,Y_i)}{\sigma_{X_i}\sigma_{Y_i}}\\&=\sum_{i\in\mathbf{R}_{X,Y}}\frac{\rho_i\sigma_{X_i}\sigma_{Y_i}+\kappa\tau_X\tau_Y}{\sigma_{X_i}\sigma_{Y_i}}=\sum_{i\in\mathbf{R}_{X,Y}}\rho_i+\kappa\tau_X\tau_Y\sum_{i\in\mathbf{R}_{X,Y}}\frac{1}{\sigma_{X_i}\sigma_{Y_i}}\end{aligned}$$

The expectation of the second summation is

$$\mathrm{Cov}(\bar{X}_2,\bar{Y}_2)\sum_{i\in\mathbf{R}_{X,Y}}\frac{1}{\sigma_{X_i}\sigma_{Y_i}}$$

Writing both $\bar{X}_2$ and $\bar{Y}_2$ in terms of their weighted sums, expanding $\mathrm{Cov}(\bar{X}_2,\bar{Y}_2)$ in terms of the covariances $\mathrm{Cov}(X_i, Y_j)$, and again noting only the covariance where $i=j$ is non-zero, the expectation of the second summation is

$$\frac{1}{\sum_{i\in\mathbf{R}_{X,Y}}\frac{1}{\sigma_{X_i}\sigma_{Y_i}}}\sum_{i\in\mathbf{R}_{X,Y}}\mathrm{Cov}\left(\frac{X_i}{\sigma_{X_i}\sigma_{Y_i}},\frac{Y_i}{\sigma_{X_i}\sigma_{Y_i}}\right)=\frac{\sum_{i\in\mathbf{R}_{X,Y}}\frac{\rho_i}{\sigma_{X_i}\sigma_{Y_i}}}{\sum_{i\in\mathbf{R}_{X,Y}}\frac{1}{\sigma_{X_i}\sigma_{Y_i}}}+\frac{\kappa\tau_X\tau_Y\sum_{i\in\mathbf{R}_{X,Y}}\frac{1}{\sigma^2_{X_i}\sigma^2_{Y_i}}}{\sum_{i\in\mathbf{R}_{X,Y}}\frac{1}{\sigma_{X_i}\sigma_{Y_i}}}$$

Subtracting the second of these expectations from the first gives (9), where a and b are given by (10) and (11), respectively.

Appendix B: Deriving the Expectation $e_{(1,2)}=e_{(2,1)}$ for Meta-Regression

Consider the following linear transformations when fitting the models required when computing (16): $\mathbf{U}_X=\Psi^{1/2}\mathbf{X}\sim N(\mathbf{V}_X\beta_X=\Psi^{1/2}\mathbf{W}_X\beta_X,\mathbf{I})$ and $\mathbf{U}_Y=\Psi^{1/2}\mathbf{Y}\sim N(\mathbf{V}_Y\beta_Y=\Psi^{1/2}\mathbf{W}_Y\beta_Y,\mathbf{I})$, where $\mathbf{I}$ denotes the $n_{X,Y}$ by $n_{X,Y}$ identity matrix. The regression estimates in (16) are unchanged by these transformations and $q_{(1,2)}=\hat{\boldsymbol{\varepsilon}}_X^{\mathrm{T}}\hat{\boldsymbol{\varepsilon}}_Y$ where $\hat{\boldsymbol{\varepsilon}}_X$ are the residuals from the fitted regression of $\mathbf{U}_X$ on $\mathbf{V}_X$, and $\hat{\boldsymbol{\varepsilon}}_Y$ are the residuals from the fitted regression of $\mathbf{U}_Y$ on $\mathbf{V}_Y$. As these fitted models are unweighted regressions we have

the usual

$$\begin{aligned}\hat{\boldsymbol{\varepsilon}}_X &= [\mathbf{I}-\mathbf{V}_X(\mathbf{V}_X^{\mathrm{T}}\mathbf{V}_X)^{-1}\mathbf{V}_X^{\mathrm{T}}]\mathbf{U}_X \\ &= [\mathbf{I}-\mathbf{V}_X(\mathbf{V}_X^{\mathrm{T}}\mathbf{V}_X)^{-1}\mathbf{V}_X^{\mathrm{T}}](\mathbf{V}_X\beta_X+\boldsymbol{\varepsilon}_X) \\ &= [\mathbf{I}-\mathbf{V}_X(\mathbf{V}_X^{\mathrm{T}}\mathbf{V}_X)^{-1}\mathbf{V}_X^{\mathrm{T}}]\boldsymbol{\varepsilon}_X = \mathbf{A}_X\boldsymbol{\varepsilon}_X\end{aligned}$$

and similarly

$$\hat{\boldsymbol{\varepsilon}}_Y = [\mathbf{I}-\mathbf{V}_Y(\mathbf{V}_Y^{\mathrm{T}}\mathbf{V}_Y)^{-1}\mathbf{V}_Y^{\mathrm{T}}]\boldsymbol{\varepsilon}_Y = \mathbf{A}_Y\boldsymbol{\varepsilon}_Y$$

where, from model (15), $\boldsymbol{\varepsilon}_X$ and $\boldsymbol{\varepsilon}_Y$ are jointly multivariate $2n$ normal with expectation $\mathbf{0}$; it should further be noted that

$$\mathrm{Cov}(\boldsymbol{\varepsilon}_{X_i}, \boldsymbol{\varepsilon}_{Y_i}) = \rho_i + \frac{\kappa\tau_X\tau_Y}{\sigma_{X_i}\sigma_{Y_i}} \tag{B1}$$

and

$$\mathrm{Cov}(\boldsymbol{\varepsilon}_{X_i}, \boldsymbol{\varepsilon}_{Y_j}) = 0 \tag{B2}$$

for $i \neq j$. Write $\mathbf{A} = \mathbf{A}_X^{\mathrm{T}}\mathbf{A}_Y$, so that $q_{(1,2)} = \boldsymbol{\varepsilon}_X^{\mathrm{T}}\mathbf{A}\boldsymbol{\varepsilon}_Y$ and

$$q_{(1,2)} = \mathrm{tr}(q_{(1,2)}) = \mathrm{tr}(\boldsymbol{\varepsilon}_X^{\mathrm{T}}\mathbf{A}\boldsymbol{\varepsilon}_Y) = \mathrm{tr}(\mathbf{A}\boldsymbol{\varepsilon}_Y\boldsymbol{\varepsilon}_X^{\mathrm{T}})$$

Interchanging the trace and expectation operations gives

$$e_{(1,2)} = \mathrm{tr}(\mathbf{A}E[\boldsymbol{\varepsilon}_Y\boldsymbol{\varepsilon}_X^{\mathrm{T}}]) = \mathrm{tr}(\mathbf{AC})$$

where $\mathbf{C}$ denotes $\mathrm{Cov}(\boldsymbol{\varepsilon}_Y, \boldsymbol{\varepsilon}_X^{\mathrm{T}})$. From Equations (B1) and (B2) we have that $\mathbf{C} = \mathbf{P} + \kappa\tau_1\tau_2\Psi$, where $\mathbf{P} = \mathrm{diag}(\rho_i)$, so that $e_{(1,2)} = \mathrm{tr}(\mathbf{A}(\mathbf{P} + \kappa\tau_1\tau_2\Psi))$ giving (17)–(19).

Acknowledgements

The authors are employed by the U.K. Medical Research Council (grant codes U.1052.00.006 and U.1052.00.001).

References

1. Sutton AJ, Abrams KR, Jones D, Sheldon DR, Song F. *Methods for Meta-analysis in Medical Research*. Wiley: New York, 2002.
2. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* 1999; **18**:2693–2708.
3. Higgins JPT, Thompson SG. Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine* 2004; **23**:1663–1682.
4. Riley RD, Abrams KR, Lambert PC, Sutton AJ, Thompson JR. An evaluation of bivariate random effects meta-analysis for the joint synthesis of two correlated outcomes. *Statistics in Medicine* 2007; **26**:78–97.
5. Riley RD, Abrams KR, Lambert PC, Sutton AJ, Thompson JR. Bivariate random effects meta-analysis and the estimation of between-study correlation. *BMC Medical Research Methodology* 2007; **7**(3).
6. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine* 2002; **21**:589–624.
7. Jackson D, Copas J, Sutton AJ. Modelling reporting bias: the operative mortality rate for ruptured abdominal aortic aneurysm repair. *Journal of the Royal Statistical Society, Series A* 2005; **168**:737–752.
8. Thompson SG, Pocock SJ. Can meta-analyses be trusted? *Lancet* 2008; **338**:1464–1465.
9. Jackson D. The significance level of the standard test for a treatment effect in meta-analysis. *Statistics in Biopharmaceutical Research* 2009; **1**:92–100.
10. Biggerstaff BJ, Tweedie RL. Incorporating variability of estimates of heterogeneity in the random effects model in meta-analysis. *Statistics in Medicine* 1997; **16**:753–768.
11. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**:177–188.
12. Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. *Statistics in Medicine* 1996; **15**:619–629.
13. Normand SLT. Meta-analysis: formulating, evaluating, combining and reporting. *Statistics in Medicine* 1999; **18**:321–359.
14. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 1997; **53**:983–997.
15. Pinto EM, Willan AR, O'Brien BJ. Cost Ŭeffectiveness analysis for multinational clinical trials. *Statistics in Medicine* 1997; **24**:1965–1982.
16. Olkin I, Trikalinos TA. A method for the meta-analysis of mutually exclusive binary outcomes. *Statistics in Medicine* 2008; **27**:4279–4300.
17. Riley RD, Thompson JR, Abrams KR. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics* 2008; **9**:172–186.
18. Ishak KJ, Platt RW, Joseph L, Hanley JA. Impact of approximating or ignoring within-study covariances in multivariate meta-analyses. *Statistics in Medicine* 2008; **27**:670–686.
19. Jennrich RI, Schluchter MD. Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* 1986; **42**:805–820.

20. Hartung J, Knapp G. On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Statistics in Medicine* 2001; **20**:1771–1782.
21. Hartung J, Knapp G. A refined method for the meta-analysis of controlled clinical trials with binary outcomes. *Statistics in Medicine* 2001; **20**:3875–3889.
22. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine* 2003; **22**:2693–2710.
23. Sohn SY. Multivariate meta-analysis with potentially correlated marketing study results. *Naval Research Logistics* 2000; **47**:500–510.
24. Li CK. A simple proof of the Craig–Sakamoto theorem. *Linear Algebra and its Applications* 2000; **321**:281–283.
25. Brockwell SE, Gordon IR. A simple method for inference on an overall effect in meta-analysis. *Statistics in Medicine* 2007; **26**:4531–4543.
26. Higgins JPT, Thompson SG. Quantifying heterogeneity in meta-analysis. *Statistics in Medicine* 2002; **21**:1539–1558.
27. Follmann DA, Proschan MA. Valid inference in random effects meta-analysis. *Biometrics* 1999; **55**:732–737.
28. Sidik K, Jonkman J. A comparison of heterogeneity variance estimators in combining results of studies. *Statistics in Medicine* 2007; **26**:1964–1981.
29. White IR. Multivariate random effects meta-analysis. *The Stata Journal* 2009; **9**:40–56.
30. Fibrinogen Studies Collaboration. Collaborative meta-analysis of prospective studies of plasma fibrinogen and cardiovascular disease. *European Journal of Cardiovascular Prevention and Rehabilitation* 2004; **11**:9–16.
31. Fibrinogen Studies Collaboration. Plasma fibrinogen and the risk of major cardiovascular diseases and non-vascular mortality: meta-analysis of individual data for 154 211 adults in 31 prospective studies. *Journal of the American Medical Association* 2005; **294**:1799–1809.
32. DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. *Contemporary Clinical Trials* 2007; **28**:105–114.
33. Sidik K, Jonkman J. Simple heterogeneity variance for meta-analysis. *Journal of the Royal Statistical Society, Series C* 2005; **54**:367–384.
34. Hartung J, Makambi KH. Positive estimation of the between-group variance component in one-way ANOVA and meta-analysis. *South African Statistical Journal* 2002; **36**:55–76.
35. Taye HH, van Houwelingen HC, Stijnen T. The binomial distribution of meta-analysis was preferred to model within-study variability. *Journal of Clinical Epidemiology* 2008; **61**:41–51.
36. Lee KJ, Thompson SG. Flexible parametric models for random effects distributions. *Statistics in Medicine* 2008; **27**:418–434.
37. Baker R, Jackson D. A new approach to outliers in meta-analysis. *Health Care in Management Science* 2008; **11**:121–131.
38. Böhning D, Malzahn U, Dietz E, Schlattmann P, Viwatwongkasem C, Biggeri A. Some general points in estimating heterogeneity variance with the DerSimonian–Laird estimator. *Biostatistics* 2002; **3**:445–457.