Journal of the Royal Statistical Society
SERIES A Statistics in Society

J. R. Statist. Soc. A (2009)
172, *Part* 4, *pp.* 789–811

Multivariate meta-analysis: the effect of ignoring within-study correlation

Richard D. Riley

University of Birmingham, UK

[Received May 2008. Revised October 2008]

Summary. Multivariate meta-analysis allows the joint synthesis of summary estimates from multiple end points and accounts for their within-study and between-study correlation. Yet practitioners usually meta-analyse each end point independently. I examine the role of within-study correlation in multivariate meta-analysis, to elicit the consequences of ignoring it. Using analytic reasoning and a simulation study, the within-study correlation is shown to influence the 'borrowing of strength' across end points, and wrongly ignoring it gives meta-analysis results with generally inferior statistical properties; for example, on average it increases the mean-square error and standard error of pooled estimates, and for non-ignorable missing data it increases their bias. The influence of within-study correlation is only negligible when the within-study variation is small relative to the between-study variation, or when very small differences exist across studies in the within-study covariance matrices. The findings are demonstrated by applied examples within medicine, dentistry and education. Meta-analysts are thus encouraged to account for the correlation between end points. To facilitate this, I conclude by reviewing options for multivariate meta-analysis when within-study correlations are unknown; these include obtaining individual patient data, using external information, performing sensitivity analyses and using alternatively parameterized models.

Keywords: Bivariate random-effects meta-analysis; Multiple end points; Multiple outcomes; Multivariate meta-analysis; Unknown within-study correlation

'In the realm of research synthesis ... the consequences of accounting for (modeling) dependence or ignoring it are not well understood' (Becker *et al.*, 2004).

1. Introduction

In meta-analysis multiple pooled results are required when there are multiple end points of interest across studies, such as multiple outcomes (Berkey *et al.*, 1995), multiple time points (Dear, 1994) and multiple treatment effects (Gleser and Olkin, 1994). For example, in hypertension trials the treatment effect on both systolic and diastolic blood pressure is of interest. In this situation, a multivariate meta-analysis model allows a joint synthesis of the multiple end points (Raudenbush *et al.*, 1988; Becker, 2000; Van Houwelingen *et al.*, 2002). This produces a pooled result for each end point simultaneously and can account for any correlation between end points, which may exist both within studies and between studies. The *within-study correlation* indicates the association between the summary end point estimates within a study and is caused by, for example, each individual in a study contributing data towards each end point. The *between-study correlation* indicates how the true underlying end point summary values are

Address for correspondence: Richard D. Riley, Department of Public Health, Epidemiology and Biostatistics, School of Health Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK.
E-mail: r.d.riley@bham.ac.uk

 0964–1998/09/172789

related across studies and is caused by differences across studies in patient characteristics, such as age, or changes in study characteristics, like the threshold level in diagnostic studies.

Multivariate meta-analysis has been promoted for over 20 years, with applications in education (Hedges and Olkin, 1985; Raudenbush *et al.*, 1988; Kalaian and Raudenbush, 1996; Gleser and Olkin, 1994), dentistry (Berkey *et al.*, 1995, 1998), clinical trials (Van Houwelingen *et al.*, 1993; Berkey *et al.*, 1996; Arends *et al.*, 2003; Ishak *et al.*, 2007), survival (Dear, 1994), marketing (Sohn, 2000), surrogate outcomes (Daniels and Hughes, 1997; Gail *et al.*, 2000), prognostic markers (Riley, Abrams, Lambert, Sutton and Thompson, 2007), diagnostic tests (Reitsma *et al.*, 2005; Chu and Cole, 2006; Riley, Abrams, Sutton, Lambert and Thompson, 2007; Harbord *et al.*, 2007) and genetic epidemiology (Thompson *et al.*, 2005), among others. The approach offers several advantages over a separate univariate synthesis of each end point, which assumes that the end points are independent. For instance, it allows us to describe the multivariate relationship between end points (Van Houwelingen *et al.*, 2002), to obtain joint confidence regions (Reitsma *et al.*, 2005), to model, test or make predictions from their association (Daniels and Hughes, 1997) and to estimate some function of the pooled end points (Thompson *et al.*, 2005; Reitsma *et al.*, 2005; Riley, Abrams, Lambert, Sutton and Thompson, 2007). The pooled estimates themselves also have better statistical properties due to utilization of correlation, especially when some end points are missing at random across studies (Riley, Abrams, Lambert, Sutton and Thompson, 2007; Riley, Abrams, Sutton, Lambert and Thompson, 2007).

Despite these advantages, multivariate meta-analysis methods are rarely used by practitioners in systematic reviews. The main exception is in systematic reviews of diagnostic test studies, where the bivariate random-effects meta-analysis (BRMA) approach for jointly synthesizing sensitivity and specificity is increasingly used (Reitsma *et al.*, 2005; Chu and Cole, 2006; Riley, Abrams, Sutton, Lambert and Thompson, 2007; Harbord *et al.*, 2007). In other multiple endpoint situations, practitioners usually choose to perform a separate univariate meta-analysis of each end point. The reasons for this may include tradition, the increased complexity of the multivariate approach, the need for specialist statistical software and a lack of understanding the consequences of ignoring correlation in meta-analysis, pertaining to the quote by Becker *et al.* (2004).

Another major stumbling-block is that multivariate meta-analysis models require, from each study, the within-study correlation between the summary estimates for each pair of end points. This adds to the standard requirement in a univariate meta-analysis for summary (effect) estimates and their standard errors to be obtained for each study, which is not necessarily trivial itself. In some scenarios the within-study correlations can justifiably be assumed zero or close to zero (Reitsma *et al.*, 2005; Daniels and Hughes, 1997; Korn *et al.*, 2005; Thompson *et al.*, 2005; Van Houwelingen *et al.*, 2002), such as in diagnostic studies where sensitivity and specificity estimates are independently derived from separate patients. However, in other settings such as the meta-analysis of longitudinal data (Ishak *et al.*, 2007), or for multiple outcomes such as overall and disease-free survival (Riley, Abrams, Lambert, Sutton and Thompson, 2007), the true within-study correlations are likely to be non-zero. Unfortunately, within-study correlations are rarely reported in primary study publications, and conceivably this must put practitioners off the multivariate meta-analysis approach. However, the Campbell Collaboration states that meta-analysts 'should not ignore the dependence among study outcomes' and 'should use some procedure to deal with dependence' (Becker *et al.*, 2004).

In this paper I examine the role of within-study correlation in multivariate meta-analysis and review approaches proposed for dealing with unknown within-study correlations, to facilitate both understanding and application of the multivariate approach. I focus particularly

on how within-study correlation influences pooled estimates from a BRMA, and the effect of ignoring it. Ishak *et al.* (2008) recommend that if interest lies only in pooled estimates from a BRMA then 'one can even ignore within-study correlations and assume independence without any significant risk of bias or loss of precision in estimates'. Here, I show that this conclusion is not generally true and in many situations the within-study correlations are influential towards the pooled estimates. In Section 2 I introduce a general model for BRMA, and the role of within-study correlation is then studied through analytic reasoning (Section 3), a simulation study (Section 4) and applied examples (Section 5). In Section 6 I review possible options to deal with unavailable within-study correlation, and Section 7 concludes with some discussion.

2. Bivariate random-effects meta-analysis

In this section I review the specification and estimation of a general model for BRMA and also provide some motivating examples.

2.1. Model specification

In a general model for BRMA (Van Houwelingen *et al.*, 2002), the two correlated end points of interest are assumed to follow a bivariate normal distribution both within studies and between studies and can be specified by using similar notation to that of Ishak *et al.* (2008) as follows:

$$\begin{aligned} \begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} &\sim N\left\{ \begin{pmatrix} \mu_{1i} \\ \mu_{2i} \end{pmatrix}, \mathbf{S}_i \right\}, & \mathbf{S}_i &= \begin{pmatrix} S_{11i}^2 & S_{11i}S_{22i}\rho_{S_i} \\ S_{11i}S_{22i}\rho_{S_i} & S_{22i}^2 \end{pmatrix}, \\ \begin{pmatrix} \mu_{1i} \\ \mu_{2i} \end{pmatrix} &\sim N\left\{ \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \mathbf{D} \right\}, & \mathbf{D} &= \begin{pmatrix} D_{11}^2 & D_{11}D_{22}\rho_D \\ D_{11}D_{22}\rho_D & D_{22}^2 \end{pmatrix}. \end{aligned} \tag{2.1}$$

The two end points are denoted by 1 and 2, and $i = 1, \ldots, n$ studies are assumed. As in univariate meta-analysis, each study supplies summary measures y_{1i} and y_{2i}, and associated standard errors S_{11i} and S_{22i}, which are assumed known. For instance, y_{1i} and y_{2i} may be the log-hazard-ratio for disease-free survival and overall survival (Riley, Abrams, Lambert, Sutton and Thompson, 2007), or the treatment effect on systolic and diastolic blood pressure (Riley, Lambert, Staessen, Wang, Gueyffier, Thijs and Boutitie, 2008). Also required is their within-study correlation ρ_{S_i} for each study (or alternatively the within-study covariance $S_{12i} = \rho_{S_i} S_{11i} S_{22i}$), which is also assumed known. Model (2.1) can accommodate studies where y_{1i} or y_{2i} are missing, under a missingness at random assumption, and for these studies the within-study correlation is not required as only one end point is included. The summary measures in each study are assumed drawn from a bivariate normal distribution with means μ_{1i} and μ_{2i}, which in turn are assumed drawn from a bivariate normal distribution with mean, or 'pooled', values of θ_1 and θ_2 respectively. The between-study variances D_{11}^2 and D_{22}^2 account for any heterogeneity in μ_{1i} and μ_{2i} across studies, and ρ_D represents their between-study correlation. Model (2.1) reverts to a bivariate fixed effects meta-analysis when $\mathbf{D} = \mathbf{0}$, and to a separate univariate meta-analysis of each end point when $\rho_{S_i} = \rho_D = 0$, i.e. all correlations are 0.

2.2. Estimation

Usually of key interest from model (2.1) are the pooled end point estimates $\hat{\theta}_1$ and $\hat{\theta}_2$, and their associated uncertainty, perhaps with a joint confidence or prediction interval (Ades *et al.*, 2005; Reitsma *et al.*, 2005; Higgins *et al.*, 2008). $\hat{\theta}_1$ and $\hat{\theta}_2$ can be obtained iteratively alongside

estimation of the between-study covariance matrix. In a classical framework, this can be achieved by using restricted iterative generalized least squares or equivalently restricted maximum likelihood, as described elsewhere (Van Houwelingen *et al.*, 2002). In terms of the pooled estimates this produces, at each iteration, the solutions

$$\hat{\boldsymbol{\theta}} = \left\{ \sum_{i=1}^{n} (\mathbf{S}_i + \hat{\mathbf{D}})^{-1} \right\}^{-1} \sum_{i=1}^{n} (\mathbf{S}_i + \hat{\mathbf{D}})^{-1} \mathbf{Y}, \tag{2.2}$$

$$\text{cov}(\hat{\boldsymbol{\theta}}) = \left\{ \sum_{i=1}^{n} (\mathbf{S}_i + \hat{\mathbf{D}})^{-1} \right\}^{-1}. \tag{2.3}$$

Here $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2)$, $\mathbf{Y}$ is a column vector of the summary measures from each study, which is formed by stacking $\mathbf{Y}_i = (y_{1i}, y_{2i})'$ from $i = 1, \ldots, n$, and $\hat{\mathbf{D}}$ is the estimate of the between-study covariance matrix from the previous iteration. Cholesky decomposition (Gentle, 1998)

Table 1. Periodontal data of Berkey *et al.* (1995,1998), providing the mean difference between a surgical and non-surgical procedure for treating periodontal disease, with improvement in probing depth and improvement in attachment level the two end points of interest (measured in millimetres 1 year after treatment)

Study	*Mean improvement in probing depth (surgical – non-surgical)*		*Mean improvement in attachment level (surgical – non-surgical)*		S_{12i}	ρ_{S_i}
	Y_{1i}	S^2_{11i}	Y_{2i}	S^2_{22i}		
1	0.47	0.0075	−0.32	0.0077	0.0030	0.39
2	0.20	0.0057	−0.60	0.0008	0.0009	0.42
3	0.40	0.0021	−0.12	0.0014	0.0007	0.41
4	0.26	0.0029	−0.31	0.0015	0.0009	0.43
5	0.56	0.0148	−0.39	0.0304	0.0072	0.34

Table 2. Scholastic aptitude test data of Gleser and Olkin (1994), providing the standardized mean difference between coached students and non-coached students for each of mathematics and verbal scholastic aptitude test scores

Study	*Mean standardized mathematics scholastic aptitude test score (coached – uncoached)*		*Mean standardized verbal scholastic aptitude test score (coached – uncoached)*		S_{12i}	ρ_{S_i}
	Y_{1i}	S^2_{11i}	Y_{2i}	S^2_{22i}		
1	1.189	0.09	0.608	0.08	0.08	0.648
2	0.652	0.128	−0.148	0.122	0.122	0.632
3	−0.065	0.038	0.124	0.039	0.039	0.649
4	−0.078	0.148	0.403	0.151	0.151	0.656
5	0.373	0.033	−0.246	0.032	0.032	0.646
6	0.186	0.045	0.137	0.045	0.045	0.644
7	−0.951	0.278	0.61	0.262	0.262	0.567

Table 3. Tumour marker data of Riley *et al.* (2003), providing the ln(hazard ratio) for whether amplified levels of MYCN and deletion of chromosome 1p are associated with a worse disease-free survival in children with neuroblastoma†

Study	*ln(hazard ratio) for MYCN (amplified versus non-amplified)*		*ln(hazard ratio) for chromosome 1p (deletion versus non-deletion)*		S_{12i}	ρ_{S_i}
	Y_{1i}	S^2_{11i}	Y_{2i}	S^2_{22i}		
1	1.64	0.26	NA	NA	—	—
2	−0.11	0.45	NA	NA	—	—
3	1.46	0.17	0.80	0.19	NA	NA
4	1.64	0.41	NA	NA	—	—
5	2.04	0.38	NA	NA	—	—
6	2.19	0.18	NA	NA	—	—
7	1.34	0.26	NA	NA	—	—
8	0.55	0.15	NA	NA	—	—
9	0.93	0.11	NA	NA	—	—
10	1.52	0.12	NA	NA	—	—
11	0.41	0.67	0.65	0.71	NA	NA
12	1.44	1.37	NA	NA	—	—
13	2.19	0.12	3.06	0.35		
14	1.06	0.29	NA	NA	—	—
15	1.18	0.33	NA	NA	—	—
16	1.85	0.44	NA	NA	—	—
17	1.77	0.21	NA	NA	—	—
18	2.95	1.18	NA	NA	—	—
19	2.98	0.33	NA	NA	—	—
20	1.90	0.21	1.67	0.21	NA	NA
21	2.50	0.58	NA	NA	—	—
22	1.92	0.12	1.90	0.12	NA	NA
23	1.90	0.78	NA	NA	—	—
24	2.39	0.53	0.17	0.40	NA	NA
25	0.47	0.28	NA	NA	—	—
26	3.29	0.25	NA	NA	—	—
27	1.90	0.33	NA	NA	—	—
28	1.62	0.17	NA	NA	—	—
29	2.56	0.30	NA	NA	—	—
30	1.70	0.15	NA	NA	—	—
31	0.84	0.07	NA	NA	—	—
32	1.60	0.24	2.10	1.23	NA	NA
33	1.64	0.41	NA	NA	—	—
34	1.45	0.33	NA	NA	—	—
35	NA	NA	1.48	0.20	NA	NA
36	0.25	0.08	NA	NA	—	—
37	2.37	1.00	NA	NA	—	—
38	0.52	0.17	NA	NA	—	—
39	1.43	0.14	NA	NA	—	—
40	0.30	0.07	0.79	0.04	NA	NA
41	0.29	0.35	NA	NA	—	—
42	0.76	0.24	NA	NA	—	—
43	5.70	3.00	NA	NA	—	—

†NA, not available.

of $\mathbf{D}$ is often required to ensure that $\hat{\mathbf{D}}$ is estimated positive semidefinite and therefore that the between-study correlation estimate $\hat{\rho}_D$ is in the range $[-1,1]$ (Van Houwelingen *et al.*, 2002; Riley, Abrams, Sutton, Lambert and Thompson, 2007). A Bayesian approach to model (2.1) is also possible (Nam *et al.*, 2003), additionally requiring specification of prior distributions for $\boldsymbol{\theta}$ and $\hat{\mathbf{D}}$, with estimation of the posterior distribution then achieved by using Markov chain Monte Carlo simulation, e.g. in WinBUGS (Lunn *et al.*, 2000). The Bayesian approach is discussed further in Section 6.4, but I primarily focus on classical estimation in this paper.

2.3. Motivating data sets

To motivate the application of BRMA, in Tables 1–3 I present three meta-analysis data sets from dental, educational and medical research. Each data set contains two end points whose summary estimates are potentially correlated, and they cover both complete-data and missing data scenarios. The two complete-data examples involve known within-study correlations, whereas the missing data example has the more common problem of unknown within-study correlations. Application to each data set will be made in Section 5.

3. Analytic assessment of the effect of within-study correlation on pooled estimates

As discussed, estimation of model (2.1) requires ρ_{S_i} (or equivalently S_{12i}) to be available in those studies providing both end points, which is unfortunately unlikely. For this reason Ishak *et al.* (2008) examine, in a variety of simulated settings, how restricted maximum likelihood estimates from model (2.1) perform when the within-study correlations are truly known compared with when they are approximated or ignored (i.e. set to 0). They conclude that

> '(pooled) effects or heterogeneity parameters were generally similar in both meta-analyses based on observed or approximated within-study covariances',

and thus if interest lies only in the pooled effects we can fit model (2.1) with S_{12i} set to 0 'without any significant risk of bias or loss of precision in estimates'. In this section, I examine this recommendation further by analytically assessing the influence of within-study correlation on the pooled estimates.

3.1. The size of within-study variation relative to between-study variation

In the BRMA of model (2.1), the μ_{ji} are often nuisance parameters as they are rarely of interest. Inference and estimation are thus usually based on the marginal model, which is written as

$$\begin{gathered}\begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} \sim N\left\{ \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \mathbf{V}_i \right\}, \\ \mathbf{V}_i = \mathbf{S}_i + \mathbf{D} = \begin{pmatrix} S_{11i}^2 + D_{11}^2 & S_{11i}S_{22i}\rho_{S_i} + D_{11}D_{22}\rho_D \\ S_{11i}S_{22i}\rho_{S_i} + D_{11}D_{22}\rho_D & S_{22i}^2 + D_{22}^2 \end{pmatrix}. \end{gathered} \tag{3.1}$$

The diagonal terms in $\mathbf{V}_i$ indicate that the total variability affecting the summary measures in each study is the sum of the within-study variance and the between-study variance. Similarly, the off-diagonal terms in $\mathbf{V}_i$ indicate that the total covariance that is contributed by each study is the sum of the within-study covariance and the between-study covariance. Equations (2.2) and (2.3) show that the total variance and total covariance are both important in the estimation of the pooled effects and their precision. Now, the influence of the within-study correlation on the total covariance depends on the size of the within-study variation relative to the between-study

variation. As the relative size of the within-study variation becomes increasingly large, $\mathbf{V}_i \to \mathbf{S}_i$ and the within-study correlations are thus potentially very important. This situation relates to when a bivariate fixed effects meta-analysis is appropriate, and random effects are not necessary. Conversely, as the relative size of the within-study variation becomes increasingly small, $\mathbf{V}_i \to \mathbf{D}_i$ and so the between-study correlation is now important, but the size of within-study correlation is irrelevant. Thus, in situations where the between-study variation dominates every $\mathbf{V}_i$, the within-study correlations will have little effect regardless of their value. An indicator of such situations is when both end points have a large I^2-value close to 100%. (Higgins *et al.*, 2003), as this indicates that nearly all the total variation for each end point is due to between-study heterogeneity.

3.1.1. Link to the Ishak simulations

All the Ishak *et al.* (2008) simulations contain small within-study variation relative to the between-study variation, i.e. S^2_{11i} and S^2_{22i} are consistently small relative to D^2_{11} and D^2_{22}. For example, their scenario 2b involves the smallest number of patients with an average study size of 33 patients which, using the formula in section 3.3 of Ishak *et al.* (2008), relates to $S^2_{11i} \approx 0.30$ and $S^2_{22i} \approx 0.15$; these are much smaller than the chosen between-study variances of $D^2_{11} = 5.0$ and $D^2_{22} = 2.5$. In such situations $\mathbf{V}_i \to \mathbf{D}_i$ and we would not expect the within-study correlations to have much effect, as ultimately borne out in the simulation results of Ishak *et al.* (2008). Sohn (2000) has also shown that even the within-study variances S^2_{11i} and S^2_{22i} have little influence in such situations. The question thus remains regarding the exact influence of the within-study correlation on the pooled estimates in other situations, e.g. where the within-study variation is at least similar in size to the between-study variation.

3.2. Role of within-study correlation in 'borrowing strength'

Riley, Abrams, Lambert, Sutton and Thompson (2007) and Riley, Abrams, Sutton, Lambert and Thompson (2007) showed that a BRMA utilizes correlation to 'borrow strength' across end points, leading to pooled estimates with smaller standard error and mean-square error than those from univariate random-effects meta-analysis (URMA) on average. Thus to understand fully the effect of ignoring within-study correlation, its role in borrowing strength needs to be understood. This can be achieved by studying the analytic solutions for the pooled estimates at each iteration. Consider the pooled estimate for end point 1, which can be written analytically at each iteration as (Riley, 2005)

$$\hat{\theta}_1 = \sum_{i=1}^{n} a_i y_{1i} + b_i y_{2i} \tag{3.2}$$

where

$$\begin{aligned} a_i = {} & \frac{1}{(\hat{D}^2_{11}+S^2_{11i})(\hat{D}^2_{22}+S^2_{22i})-(\hat{D}_{12}+S_{12i})^2} \sum_{k=1}^{n} \frac{(\hat{D}^2_{22}+S^2_{22i})(\hat{D}^2_{11}+S^2_{11k})-(\hat{D}_{12}+S_{12i})(\hat{D}_{12}+S_{12k})}{(\hat{D}^2_{11}+S^2_{11k})(\hat{D}^2_{22}+S^2_{22k})-(\hat{D}_{12}+S_{12k})^2} \\ & \times \left[\sum_{k=1}^{n} \frac{\hat{D}^2_{11}+S^2_{11k}}{(\hat{D}^2_{11}+S^2_{11k})(\hat{D}^2_{22}+S^2_{22k})-(\hat{D}_{12}+S_{12k})^2} \sum_{k=1}^{n} \frac{\hat{D}^2_{22}+S^2_{22k}}{(\hat{D}^2_{11}+S^2_{11k})(\hat{D}^2_{22}+S^2_{22k})-(\hat{D}_{12}+S_{12k})^2} \right. \\ & \left. - \left\{ \sum_{k=1}^{n} \frac{(\hat{D}_{12}+S_{12k})}{(\hat{D}^2_{11}+S^2_{11k})(\hat{D}^2_{22}+S^2_{22k})-(\hat{D}_{12}+S_{12k})^2} \right\}^2 \right]^{-1}, \end{aligned} \tag{3.3}$$

$$b_i = \frac{1}{(\hat{D}_{11}^2+S_{11i}^2)(\hat{D}_{22}^2+S_{22i}^2)-(\hat{D}_{12}+S_{12i})^2}\sum_{k=1}^{n}\frac{S_{11i}^2(\hat{D}_{12}+S_{12k})-S_{11k}^2(\hat{D}_{12}+S_{12i})+\hat{D}_{11}^2(S_{12k}-S_{12i})}{(\hat{D}_{11}^2+S_{11k}^2)(\hat{D}_{22}^2+S_{22k}^2)-(\hat{D}_{12}+S_{12k})^2}$$
$$\times\left[\sum_{k=1}^{n}\frac{\hat{D}_{11}^2+S_{11k}^2}{(\hat{D}_{11}^2+S_{11k}^2)(\hat{D}_{22}^2+S_{22k}^2)-(\hat{D}_{12}+S_{12k})^2}\sum_{k=1}^{n}\frac{\hat{D}_{22}^2+S_{22k}^2}{(\hat{D}_{11}^2+S_{11k}^2)(\hat{D}_{22}^2+S_{22k}^2)-(\hat{D}_{12}+S_{12k})^2}\right.$$
$$\left.-\left\{\sum_{k=1}^{n}\frac{(\hat{D}_{12}+S_{12k})}{(\hat{D}_{11}^2+S_{11k}^2)(\hat{D}_{22}^2+S_{22k}^2)-(\hat{D}_{12}+S_{12k})^2}\right\}^2\right]^{-1}. \tag{3.4}$$

Here both i and k represent the $1,\ldots,n$ studies, with k needed to distinguish the summation from 1 to n within the summation for $i=1,\ldots,n$. Note also that $\hat{D}_{11}^2$, $\hat{D}_{22}^2$ and $\hat{D}_{12}$ are the values within the restricted maximum likelihood estimate of the between-study covariance matrix $\hat{\mathbf{D}}$ from the previous iteration. Each study thus contributes $a_i y_{1i} + b_i y_{2i}$ towards the pooled estimate $\hat{\theta}_1$. It is shown elsewhere (Riley, 2005) that $\Sigma_{i=1}^{n}(a_i+b_i)=1$ and thus each study's weighting to the value of $\hat{\theta}_1$ is a_i+b_i, or $100(a_i+b_i)\%$. The b_i-term indicates that the summary measures for end point 2 are utilized in the pooled effect estimate for end point 1. The size of b_i, and thus the degree of borrowing strength from end point 2, depends on various factors, including the size of within-study and between-study correlation. For example, when $\hat{D}_{12}=0$ and $S_{12i}=0$ for all studies, the weights reduce to those known for a URMA of

$$a_i = \frac{1/(\hat{D}_{11}^2+S_{11i}^2)}{\sum_{i=1}^{n}1/(\hat{D}_{11}^2+S_{11i}^2)} = \frac{w_i}{\sum_{i=1}^{n}w_i} \qquad b_i = 0 \tag{3.5}$$

and there is thus no borrowing of strength from end point 2. In other situations involving non-zero correlations there is more potential for borrowing strength, though the degree to which this occurs depends on the differences in the within-study covariance matrices across studies. Riley, Abrams, Lambert, Sutton and Thompson (2007) discussed that, when the $\mathbf{S}_i$ are exactly the same across studies, each b_i is 0 and the BRMA again reduces to a URMA, regardless of the size of within-study and between-study correlations. However, as the differences between the $\mathbf{S}_i$ increase, there is more opportunity for the correlations to utilize the related end point and to borrow strength. For example, the numerator of b_i in equation (3.4) contains the terms

$$S_{11i}^2(\hat{D}_{12}+S_{12k})-S_{11k}^2(\hat{D}_{12}+S_{12i})+\hat{D}_{11}^2(S_{12k}-S_{12i}). \tag{3.6}$$

This reveals that the difference across studies in S_{11i}^2 and S_{11k}^2 (the within-study variances for end point 1) and the differences in S_{12i} and S_{12k} (the within-study covariances) influence b_i, and thus the amount $\hat{\theta}_1$ borrows strength from end point 2. As $S_{12i}=\rho_{S_i}S_{11i}S_{22i}$, when the within-study correlations are unknown, simply setting them to 0 will ensure that the differences in S_{12i} and S_{12k} are zero; this will modify b_i and reduce the borrowing of strength, causing potentially pooled estimates and precision that are different from those of a BRMA correctly utilizing within-study correlation.

This is now demonstrated by a simple hypothetical example involving missing data. Missing data scenarios have the greatest potential for borrowing strength because the difference in S_{11i}^2 between known and unknown summary measures is essentially ∞ (Riley, Abrams, Lambert, Sutton and Thompson, 2007), thus maximizing the potential for large b_i. Consider that there

are only two studies in the meta-analysis, with end point 1 missing in study 1, and assume that the between-study covariance matrix is known. In this situation we can use equations (3.2)–(3.4) to find that the analytic solution for the pooled estimate for end point 1 is

$$\hat{\theta}_1 = y_{12} - \frac{(y_{22} - y_{21})(S_{112}S_{222}\rho_{S_2} + D_{11}D_{22}\rho_D)}{2D_{22}^2 + S_{221}^2 + S_{222}^2}. \tag{3.7}$$

Equation (3.7) shows that if the within-study correlation in study 2 is 0, and the between-study correlation is 0, then $\hat{\theta}_1 = y_{12}$, which is the same as the univariate meta-analysis result for end point 1. However, larger correlations and a large value of $y_{22} - y_{21}$ allow greater potential for $\hat{\theta}_1$ to move away from y_{12}. This is sensible as, when y_{22} is very different from y_{21}, and given high within- and between-study correlation, the unknown y_{11} is also likely to be very different from the known y_{12}, thus shifting $\hat{\theta}_1$ away from y_{12}. Ignoring within-study correlation here will reduce the degree of borrowing strength. For example let y_{11} be missing, $y_{21} = -1$, $y_{12} = 1$ and $y_{22} = 1$. Further, let $S_{21}^2 = S_{12}^2 = S_{22}^2 = 1$ and $D_{11}^2 = D_{22}^2 = 1$, so that the within-study and between-study variation are similar in size, and $\rho_S = \rho_D = 0.5$, such that there is moderate correlation. In this situation a URMA gives $\hat{\theta}_1 = 1$, but a BRMA uses equation (3.7) to give

$$\begin{aligned}\hat{\theta}_1 &= y_{12} - \frac{(y_{22} - y_{21})(S_{112}S_{222}\rho_{S_2} + D_{11}D_{22}\rho_D)}{2D_{22}^2 + S_{221}^2 + S_{222}^2} \\ &= 1 - \frac{\{1 - (-1)\}(0.5 + 0.5)}{4} = 0.5.\end{aligned}$$

The BRMA uses the correlation to borrow strength from end point 2, and this moves the pooled result away from the URMA solution, towards 0. Now, if the within-study correlations are wrongly assumed to be 0, the BRMA pooled result becomes $\hat{\theta}_1 = 0.75$. In contrast, if the within-study correlations are wrongly approximated as 0.9 then $\hat{\theta}_1 = 0.3$. The choice of within-study correlation thus clearly influences the pooled result here. Note though that, if the between-study variation is alternatively $D_{11}^2 = D_{22}^2 = 10$, so that it is large relative to the within-study variation, the choice of within-study correlation has less effect as discussed in Section 3.1. Using equation (3.7), the true BRMA pooled estimate for end point 1 is now $\hat{\theta}_1 = 0.5$, but assuming zero within-study correlation has only a small effect, leading to $\hat{\theta}_1 = 0.55$.

The example above contained missing data, but complete-data scenarios may somewhat mirror missing data scenarios when the number of participants differs considerably across studies, as then large differences may occur in the within-study variances. In such situations the size of the within-study correlations can again influence the pooled estimates as demonstrated above; an applied example of this is shown in Section 5.1.

4. Simulation assessment of the effect of ignoring within-study correlation

I now present results of a simulation study to contrast the statistical properties of pooled estimates from BRMA when utilizing or ignoring within-study correlations. The simulations are an extension of those described in detail elsewhere (Riley, Abrams, Sutton, Lambert and Thompson, 2007) and were performed in four different scenarios. Each scenario relates to a different specification of the BRMA model of equation (2.1), from which I generated 1000 meta-analysis data sets for subsequent analysis. All scenarios were chosen to highlight those situations that were identified in Section 3 where a BRMA can especially borrow strength, and they thus contrast the Ishak *et al.* (2008) simulations. The scenarios all involved

Table 4. Results† of a simulation study to assess the effect of ignoring within-study correlation in BRMA in a variety of settings

Model	*Number of simulations compared*	*Mean bias*		*Coverage (%)*		*Mean-square error*		*Mean standard error*		*Mean correlation between $\hat{\theta}_1$ and $\hat{\theta}_2$*
		$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	
Scenario (a): n = 50 studies; complete data for 25 studies, only end point 1 data for others with end point 2 missing completely at random; between-study variation similar in size to average within-study variation; within- and between-study correlations all 0.8										
URMA	1000	−0.004	−0.005	94.6	94.9	0.010	0.022	0.102	0.146	0
General BRMA using the correct within-study correlations	1000	−0.005	−0.007	94.6	94.6	0.009	0.015	0.099	0.119	0.644
General BRMA assuming the within-study correlations are 0	1000	−0.005	−0.005	95.2	95.7	0.010	0.016	0.101	0.131	0.471
Scenario (b): n = 10 studies; complete data for 5 studies, only end point 1 data for others with end point 2 missing completely at random; between-study variation similar in size to average within-study variation; within- and between-study correlations all 0.8										
URMA	1000	−0.007	−0.016	93.5	93.2	0.050	0.092	0.217	0.262	0
General BRMA using the correct within-study correlations	1000	−0.008	−0.014	93.4	93.3	0.050	0.071	0.212	0.225	0.674
General BRMA assuming the within-study correlations are 0	1000	−0.008	−0.014	94.8	98.0	0.050	0.072	0.223	0.251	0.423
Scenario (c): n = 50 studies; complete data for end point 1, but data for end point 2 missing if its Y_{2i} was negative (i.e. non-ignorable missing data); between-study variation similar in size to average within-study variation; within- and between-study correlations all 0.8										
URMA	1000	−0.007	0.505	94.5	0.1	0.011	0.266	0.106	0.093	0
General BRMA using the correct within-study correlations	1000	−0.021	0.314	93.8	7.6	0.011	0.108	0.101	0.084	0.481
General BRMA assuming the within-study correlations are 0	1000	−0.002	0.416	97.6	0.5	0.011	0.180	0.107	0.098	0.234
Scenario (d): n = 10 studies; complete data for end point 1, but data for end point 2 missing if its Y_{2i} was negative (i.e. non-ignorable missing data); between-study variation similar in size to average within-study variation; within- and between-study correlations all 0.8										
URMA	1000	−0.006	0.491	94.1	56.2	0.050	0.275	0.211	0.190	0
General BRMA using the correct within-study correlations	1000	−0.008	0.294	94.0	74.0	0.049	0.132	0.208	0.174	0.523
General BRMA assuming the within-study correlations are 0	1000	−0.005	0.417	94.3	69.6	0.051	0.207	0.212	0.197	0.196

†The best performing models are those with mean bias closest to 0, smallest mean standard error, smallest mean-square error and coverage closest to 95%.

(a) between-study variation similar in size to the average within-study variation,
(b) reasonably large differences in the within-study variances across studies;
(c) data missing at random for end point 2 in some studies and
(d) large within-study correlation of 0.8 in each study and also a between-study correlation of 0.8.

In addition, the scenarios differed in the number of studies in the meta-analysis ($n = 10$ or $n = 50$), and also how the missing data were generated (Rubin, 1976). The four scenarios are described in Table 4 as scenarios (a)–(d).

Each of the 1000 meta-analysis data sets that were generated in each scenario were analysed separately by fitting

(a) two separate URMAs (i.e. as equation (2.1) but assuming zero within-study and between-study correlation);
(b) the general BRMA model of equation (2.1), with the within-study correlations known and
(c) the general BRMA model of equation (2.1), but assuming that the within-study correlations were 0.

Restricted maximum likelihood estimation was used to fit each model, and the mean bias, mean standard error, mean-square error of pooled estimates, $\hat{\theta}_1$ and $\hat{\theta}_2$ were then compared across models (Table 4), together with the coverage of their 95% confidence intervals. I now summarize the key findings.

4.1. Results for scenarios (a) and (b)

In scenarios (a) and (b) all models produce approximately unbiased pooled estimates for both end points (Table 4). The BRMA using the correct within-study correlations performs better than the URMA model on average, as the pooled estimates have smaller mean standard errors and mean-square errors. Such gains are only small for end point 1, which has complete data; however, the gains are much larger for end point 2, as the missing data for this end point facilitate a large borrowing of strength from end point 1. For example, in scenario (b) the mean standard error of the pooled estimate for end point 2 is 0.262 in the URMA compared with 0.225 in the BRMA using the correct correlations. The BRMA assuming zero within-study correlation also has generally better statistical properties for the pooled estimates than those from URMA, again especially for end point 2. However, it is not as accurate as the BRMA using the known within-study correlations, as the mean standard error and mean-square error are higher. For example, in scenario (b) the mean standard error of the pooled estimate for end point 2 is 0.251 when ignoring within-study correlation and 0.225 when using the correct values. Further, the correlation between $\hat{\theta}_1$ and $\hat{\theta}_2$ is also underestimated when ignoring within-study correlation, which may considerably impact on the calculation of joint confidence or prediction regions, and the precision of any function of the two estimates.

4.2. Results for scenarios (c) and (d)

Scenarios (c) and (d) consider a special case of missing data where, after generating complete data from BRMA model (2.1), the data for end point 2 were removed if $y_{2i} < 0$. This relates to non-ignorable missingness (Rubin, 1976) and is akin to authors or journals not reporting negative results for end point $j = 2$. In this situation, the URMA gives pooled estimates for end point 2 which are upwardly biased by about 0.5 on average (Table 4). However, the BRMA

allows us to borrow strength from the data for end point 1, and considerably reduces this bias. For example, in scenario (c) the BRMA assuming zero correlation reduces the bias to 0.42, but the BRMA using the correct within-study correlations performs even better and reduces the bias to 0.32. The coverage of the 95% confidence intervals for end point 2 is also superior in the BRMA utilizing the correct within-study correlations. In scenarios like this in practice it is conceivable that, owing to the greater reduction in bias, the BRMA utilizing the correct within-study correlations may even lead to clinical or scientific conclusions regarding end point 2 that are different from a URMA or a BRMA ignoring within-study correlation. In terms of end point 1, for which there are complete data, the BRMA utilizing the correct within-study correlations has much smaller gain in mean-square error and only small or no gain in mean standard error; further in terms of bias and coverage the URMA performs at least as well (Table 4). Finally, note that the correlation between $\hat{\theta}_1$ and $\hat{\theta}_2$ is underestimated as 0.196 when ignoring within-study correlation, compared with the estimate of 0.523 when using the correct within-study correlations.

5. Applied examples

To illustrate the findings from Sections 3 and 4, I now apply BRMA to each of the four meta-analysis data sets that were introduced in Section 2.3.

5.1. Complete-data examples

The periodontal data set (Berkey *et al.*, 1995, 1998) (Table 1) contains five studies that each assess the difference in a surgical and non-surgical procedure for treating periodontal disease, with improvement in probing depth and improvement in attachment level the two outcomes of interest (measured in millimetres 1 year after treatment). There are complete data, and the within-study variances are generally much smaller than the between-study variance for each end point; for example, for the attachment level outcome they average 0.0015 and the between-study variance is estimated at about 0.033. This is reflected by the large I^2-values of 68.8% and 96.4% for probing depth and attachment level respectively, indicating that the total variation is mostly dominated by the between-study variation. This situation is thus akin to simulated scenarios of Ishak *et al.* (2008), and accordingly the within-study correlation has little effect on the BRMA pooled estimates and their standard error, which are very similar when ignoring or using the correct within-study correlations (Table 5). The biggest effect is on the between-study correlation, which is increased from 0.61 to 0.78 when ignoring within-study correlation. Pooled results from URMA are also very similar to those from BRMA, which is not surprising as there are complete data and only small differences between the within-study variances, and thus we would not expect much borrowing of strength here (Section 3.2). The analyses all indicate that the surgical procedure improves probing depth by about 0.35 mm more than the non-surgical procedure, but that the non-surgical procedure improves attachment level by 0.34 mm more than the surgical procedure.

The scholastic aptitude test data of Gleser and Olkin (1994) (Table 2) contain seven studies that assess the effect of coaching on verbal and mathematics scholastic aptitude test scores. The summary measures for each outcome are the standardized mean difference between coached students and non-coached students, with positive values indicating that coaching was effective. In contrast with the periodontal data, there are reasonably large differences between the within-study variances, which facilitate the borrowing strength across end points (Section 3.2) and explain why the BRMA and URMA results are somewhat different (Table 5). Further, although

Table 5. Meta-analysis results when fitting the various models to the four applied examples using restricted maximum likelihood†

Data set	*Model*	*Assumed value of the within-study correlations*	*Pooled value end point 1,* $\hat{\theta}_1$	*Between-study variance end point 1,* $\hat{D}_{11}^2$	*Pooled value end point 2,* $\hat{\theta}_2$	*Between-study variance end point 2,* $\hat{D}_{22}^2$	*Between-study correlation* $\hat{\rho}_D$	*Global correlation* $\hat{\rho}$
Periodontal data	URMA	—	0.361 (0.059)	0.012	−0.346 (0.089)	0.033	—	—
	General BRMA	0	0.345 (0.055)	0.010	−0.325 (0.088)	0.033	0.775	—
	General BRMA	Known	0.353 (0.059)	0.012	−0.339 (0.088)	0.033	0.609	—
	Alternative BRMA	—	0.358 (0.059)	—	−0.345 (0.087)	—	—	0.458
Scholastic aptitude test scores data	URMA	—	0.246 (0.210)	0.217	0.131 (0.127)	0.040	—	—
	General BRMA	0	0.252 (0.208)	0.212	0.127 (0.127)	0.039	0.206	—
	General BRMA	Known	0.185 (0.223)	0.275	0.144 (0.131)	0.047	−0.510	—
	Alternative BRMA	—	0.247 (0.212)	—	0.132 (0.133)	—	—	−0.035
Tumour marker data	URMA	—	1.478 (0.127)	0.386	1.369 (0.278)	0.415	—	—
	General BRMA	0	1.460 (0.124)	0.373	1.442 (0.181)	0.218	1	—
	General BRMA	0.4	1.469 (0.124)	0.370	1.368 (0.238)	0.384	0.644	—
	General BRMA	0.8	1.474 (0.125)	0.378	1.322 (0.273)	0.526	0.343	—
	Alternative BRMA	—	1.473 (0.128)	—	1.355 (0.271)	—	—	0.479

†Standard errors are given in parentheses.

there is a large I^2 of 70.2% for the mathematics outcome, for the verbal outcome I^2 is 34.1%, indicating that the within-study variation is relatively large and thus that the within-study correlations are likely to be influential in the BRMA. This is exemplified in the results for the mathematics outcome. The BRMA using the correct within-study correlations gives a pooled estimate of 0.185 in favour of coaching with a standard error of 0.223; however, the BRMA ignoring within-study correlation overestimates the effect as 0.252 and underestimates its standard error as 0.208, which is a likely consequence of the between-study variances also being underestimated. The within-study correlation is thus clearly influencing the pooled results here to an important degree.

5.2. Missing data example

The tumour marker data in Table 3 contain 43 studies that assess whether amplified levels of MYCN and deletion of chromosome 1p, Ch1p, are associated with a worse disease-free survival in children with neuroblastoma (Riley *et al.*, 2003). Eight studies provide an ln(hazard ratio) estimate with standard error for both markers, but 34 studies provide only MYCN and one study only provides Ch1p. Under a missingness at random assumption, there is an opportunity to borrow strength across markers here as deletion of Ch1p and amplification of MYCN are known to be highly positively correlated (Komuro *et al.*, 1998). There is a reasonably large I^2 for both outcomes (MYCN, 65.9%; Ch1p, 66.7%); however, there are large differences between the within-study variances and some are even large relative to the between-study variation (Table 3). The within-study correlations are thus potentially influential in borrowing strength here, but unfortunately they are unknown. A sensitivity analysis was therefore undertaken to assess BRMA results by assuming within-study correlations of 0, 0.4 and 0.8. The influence of within-study correlation is especially seen in the Ch1p-results (Table 5). The BRMA ignoring within-study correlation gives a pooled ln(hazard ratio) of 1.442, with a standard error of 0.181, relating to a pooled hazard ratio of 4.23 with 95% confidence interval of 2.97–6.03. However, the BRMA assuming within-study correlations of 0.8 gives a smaller pooled ln(hazard ratio) of 1.322, with a larger standard error of 0.273, relating to a pooled hazard ratio of 3.75 with 95% confidence interval of 2.19–6.41. Although both analyses suggest that Ch1p is an indicator of poor prognosis, the analysis utilizing within-study correlations of 0.8 indicates a much smaller effect with increased uncertainty. The BRMA results when assuming a within-study correlation of 0.4 fall in between those results when assuming that the correlations are 0 or 0.8.

6. Options for multivariate meta-analysis given unknown within-study correlation

Sections 3–5 show that the within-study correlation can be influential towards the pooled estimates. Thus, when they are unknown, simply ignoring them is not generally recommended, unless they truly are 0 or close to zero as highlighted in some previous applications (Reitsma *et al.*, 2005; Daniels and Hughes, 1997; Korn *et al.*, 2005; Thompson *et al.*, 2005; Van Houwelingen *et al.*, 2002). With this in mind, I now briefly review proposals for dealing with unknown within-study correlations in multivariate meta-analysis.

6.1. Use an approximate formula for meta-analysis of survival proportions

Dear (1994), and more recently Arends *et al.* (2008), applied a multivariate meta-analysis of survival data at multiple follow-up points, where the summary measures that were required from each study are the proportion of patients alive at each of the follow-up times. To overcome

the problem of unavailable within-study correlations between time points in each study, Dear derived a formula that retrospectively estimates them and depends only on the proportion of those surviving at each time point. The formula is only approximate, as it does not account for censoring, but a simulation study suggests that any difference from the true correlation will usually only be small.

6.2. Obtain individual patient data

Meta-analysis using individual patient data (IPD), where the raw data are obtained for each study, is often termed the 'gold standard' approach for various well-documented reasons (Stewart and Tierney, 2002). For multivariate meta-analysis, availability of IPD allows us to calculate the within-study correlation directly in each study, alleviating the reliance on reported information. For example, IPD has been used to calculate the within-study correlation between the effects of treatment on systolic and diastolic blood pressure, by modelling these two outcomes jointly in each study through a bivariate regression model (Riley, Lambert, Staessen, Wang, Gueyffier, Thijs and Boutitie, 2008). In more complex modelling situations, e.g. where two different survival outcomes are of interest, bootstrapping methods may be required to obtain the within-study correlations methods using IPD (Daniels and Hughes, 1997).

6.3. Narrow the range of possible values

One issue is that IPD may not be available in all studies (Riley, Simmonds and Look, 2007), and remaining studies that only provide aggregate data may still have unavailable within-study correlations. In this situation one solution is to use the within-study correlations that are derived from IPD studies to inform the likely value of the within-study correlation in aggregate data studies. For example, the average available within-study correlation could be imputed, or sensitivity analyses could be performed by imputing over the range of observed values. For example, using a Bayesian framework, Mc Daid *et al.* (2007) used the observed within-study correlation in their available IPD studies to produce an informative prior distribution for the missing within-study correlation in other studies.

Even without IPD studies, it may be possible to narrow the range of possible values for the unknown within-study correlations. For example, Raudenbush *et al.* (1988) used external information, as is often done in meta-analysis of cluster trials where the intraclass correlation coefficient is unavailable (Donner and Klar, 2002; Rooney and Murray, 1996). This approach was taken in the tumour marker analysis in Section 5.2, where previous research indicated that the correlations were highly likely to be positive and thus sensitivity analyses were conducted over a range of positive imputed correlations. For the special situation where multiple relative risks are to be synthesized, Berrington and Cox (2003) narrowed the range of possible values for the within-study correlation by calculating lower and upper bounds from the 2×2 tables that were available from each study.The identification of a range of correlation values has similarly helped to inform meta-analysis in other contexts (Abrams *et al.*, 2005).

6.4. Perform sensitivity analyses over the entire correlation range

Where little or no information about the within-study correlations exists, a further option is to perform sensitivity analyses by imputing correlations over the entire range of values (i.e. from -1 to 1), to assess whether and how conclusions depend on the correlation that is imputed. This type of sensitivity analysis has been used in a multivariate meta-analysis of 44 trials which evaluated the effectiveness of injectable gold, auranofin and placebo on three treatment outcomes

Table 6. Bayesian meta-analysis results for the tumour marker data set†

Model	*Prior distribution for the within-study correlations*	*Pooled value end point 1,* $\hat{\theta}_1$	*Between-study variance end point 1,* $\hat{D}_{11}^2$	*Pooled value end point 2,* $\hat{\theta}_2$	*Between-study variance end point 2,* $\hat{D}_{22}^2$	*Between-study correlation* $\hat{\rho}_D$	*Within-study correlation* $\hat{\rho}_{S_i}$
URMA	—	1.482 (0.122)	0.391	1.365 (0.381)	0.551	—	—
General BRMA	uniform(−1,1)	1.493 (0.134)	0.373	1.430 (0.206)	0.339	0.840	−0.076
General BRMA	uniform(0,1)	1.477 (0.125)	0.395	1.385 (0.251)	0.328	0.576	0.213
General BRMA	uniform(0.5,1)	1.510 (0.138)	0.499	1.430 (0.294)	0.568	0.413	0.587

†Mean posterior estimates are shown for the pooled effects, and median posterior estimates for the between-study variance and correlation parameters. Standard errors are given in parentheses.

(Berkey *et al.*, 1996). No matter what correlation values were assumed, all the analyses indicated that gold was significantly better than auranofin on all three outcomes, even though the individual trials reported no significant differences. In a Bayesian framework, Nam *et al.* (2003) took a similar approach by placing a uniform(−1,1) prior distribution on the within-study correlation and then assessed whether conclusions are robust to changes in the specification of this prior. To demonstrate this idea I applied model (2.1) to the tumour marker data in a Bayesian framework, using a product normal specification of the between-study covariance matrix (Spiegelhalter *et al.*, 2000) and the following prior distributions:

$$\left.\begin{array}{ll} \mu_{1i} \sim N(0, 0.000001), & \mu_{2i} \sim N(0, 0.000001), \\ \theta_1 \sim N(0, 0.000001), & \theta_2 \sim N(0, 0.000001), \\ D_{11} \sim N(0,1)\ I(0,), & D_{22} \sim N(0,1)\ I(0,), \\ \rho_D \sim \text{uniform}(-1,1), & \rho_{S_i}, \sim \text{uniform}(-1,1), \end{array}\right\} \tag{6.1}$$

where $I(0,)$ indicates that the distribution is truncated at 0. The analysis was then repeated with prior distributions as above but with $\rho_{S_i} \sim \text{uniform}(0,1)$, and then with $\rho_{S_i} \sim \text{uniform}(0.5,1)$. A burn-in of 10000 samples was used, with posterior inferences then taken from a further 10000 samples and the results are summarized in Table 6. All the analyses suggest that MYCN and Ch1p are of prognostic importance, but the choice of prior distribution for the within-study correlation clearly influences the posterior estimates, similarly to how the frequentist BRMA results change according to whether within-study correlations of 0, 0.4 or 0.8 are imputed (Table 5). The uniform(0.5,1) prior, for example, gives the largest between-study variance estimates and the largest standard errors. Of course, as in any Bayesian analysis, the influence of *all* prior distributions should be assessed (Spiegelhalter *et al.*, 2003), not just that for the within-study correlation, especially as even 'vague' priors may be influential when the number of studies is small (Lambert *et al.*, 2004; Browne and Draper, 2000; Gelman, 2006).

6.5. Use an alternative model that does not require the within-study correlations

An alternative model for BRMA has been proposed which does not require the within-study correlations (Riley, Thompson and Abrams, 2008). This model maintains the individual weight-

ing of each study in the analysis but includes only one *overall* correlation parameter ρ, which is a hybrid measure of the within-study and between-study correlations. This removes the need to know the within-study correlations, and the data that are required to fit the model are the same as those needed for a separate URMA of each end point, which makes it widely applicable. The alternative BRMA model can be specified as

$$\begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} \sim N\left\{ \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \boldsymbol{\Phi}_i \right\},$$
$$\boldsymbol{\Phi}_i = \begin{pmatrix} S_{11i}^2 + \psi_{11}^2 & \rho\sqrt{\{(S_{11i}^2 + \psi_{11}^2)(S_{22i}^2 + \psi_{22}^2)\}} \\ \rho\sqrt{\{(S_{11i}^2 + \psi_{11}^2)(S_{22i}^2 + \psi_{22}^2)\}} & S_{22i}^2 + \psi_{22}^2 \end{pmatrix}. \tag{6.2}$$

The parameters here are as defined for model (2.1), with additionally ψ_j^2 indicating the additional variation beyond sampling error and ρ denoting the overall correlation. The model can be fitted in Stata (StataCorp, 2007) using restricted maximum likelihood via a self-written program that is available on request. Where interest lies only in the pooled estimates, or some function of them, then, unless $\hat{\rho}$ is very close to 1 or –1, the alternative model has been shown to produce appropriate pooled estimates with little bias that

(a) are very similar to those from fitting the general BRMA model where the within-study correlations are known and
(b) have better statistical properties than those from separate URMAs, especially given missing data (Riley, Thompson and Abrams, 2008).

The alternative model also produces pooled estimates with generally better statistical properties than those from the general BRMA model ignoring within-study correlations. To demonstrate this, the alternative model was applied to the simulated data sets in scenarios (a) and (c) that were described in Section 4. The results obtained are compared with those from other models in Table 7, but just across those data sets that gave a value of $-0.95 < \hat{\rho} < 0.95$ in the alternative model, as outside the range the alternative model can have estimation issues (Riley, Thompson and Abrams, 2008). In scenario (a) the mean-square error and mean standard error for end point 2, for which there are missing data, are

(i) smaller in the alternative model than in the URMA or the general BRMA ignoring within-study correlation, but
(ii) slightly larger than in the BRMA using the known within-study correlations,

emphasizing the extra value of having the within-study correlations available. In scenario (c), where there is non-ignorable missing data for end point 2, the alternative model reduces the bias in $\hat{\theta}_2$ to about 0.32, the same as the general BRMA using the known within-study correlations, whereas the general BRMA ignoring within-study correlation reduces it to only 0.42.

The alternative BRMA model was fitted to each of the three applied examples (Table 5). Compared with the general BRMA ignoring within-study correlation, in the two complete-data examples the alternative model produces pooled estimates and standard errors that are slightly closer to those from a BRMA using the known within-study correlation. In the missing data tumour marker example, where the within-study correlations are unknown, the alternative model produces estimates that are most similar to those from the general BRMA assuming within-study correlations of 0.8; in comparison the general BRMA ignoring within-study correlation overestimates the pooled hazard ratio for Ch1p and underestimates its precision.

Table 7. Results of a simulation study to compare the alternative model with the general BRMA model ignoring within-study correlation

Model	*Number of simulations compared*	*Mean bias*		*Coverage (%)*		*Mean-square error*		*Mean standard error*		*Mean correlation between* $\hat{\theta}_1$ *and* $\hat{\theta}_2$
		$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	
Scenario (a): n = 50 studies; complete data for 25 studies, only end point 1 data for others with end point 2 missing completely at random; between-study variation similar in size to average within-study variation; within- and between-study correlations all 0.8										
URMA	988	−0.005	−0.005	94.5	94.8	0.010	0.022	0.102	0.146	0
General BRMA using the correct within-study correlations	988	−0.005	−0.007	94.3	94.5	0.010	0.015	0.099	0.120	0.644
General BRMA assuming the within-study correlations are 0	988	−0.005	−0.005	95.1	95.6	0.010	0.017	0.101	0.131	0.471
Alternative BRMA model	988	−0.005	−0.007	94.3	95.1	0.010	0.015	0.099	0.124	0.618
Scenario (c): n = 50 studies; complete data for end point 1, but data for end point 2 missing if its Y_{2i} *was negative (i.e. non-ignorable missing data); between-study variation similar in size to average within-study variation; within- and between-study correlations all 0.8*										
URMA	879	−0.001	0.507	94.2	0.0	0.012	0.269	0.106	0.093	0
General BRMA using the correct within-study correlations	879	−0.019	0.320	93.9	6.6	0.011	0.111	0.102	0.085	0.489
General BRMA assuming the within-study correlations are 0	879	0.002	0.422	94.9	0.5	0.012	0.185	0.107	0.018	0.238
Alternative BRMA model	879	−0.016	0.319	94.5	11.5	0.011	0.111	0.106	0.097	0.702

7. Discussion

> 'No reviewer should ever ignore dependence among study outcomes. Even the most simplest *ad hoc* options are better than pretending such dependence does not exist' (Becker, 2000).

Multivariate meta-analysis of multiple correlated end points offers many benefits, including the borrowing of strength across end points and the calculation of joint confidence and prediction regions. Yet practitioners usually opt for a separate univariate meta-analysis of each end point, which makes the strong assumption that end points are independent. In this paper I have shown how and why pooled estimates from a BRMA depend on the within-study correlation between end points, and that simply ignoring it can alter meta-analysis results in important ways. Section 3 showed analytically that, except when within-study variation is very small relative to between-study variation, the within-study correlation plays a crucial role in the amount of borrowing of strength; it thus impacts on the pooled estimates, their precision and their correlation, which are all crucial for making further inferences. This was exemplified in the applied examples of Section 5, with the meta-analysis results noticeably dependent on the choice of within-study correlation, except in the periodontal data example as within-study variation was relatively small. The simulations of Section 4 also showed that an analysis ignoring within-study correlation is less accurate as, on average, it increases the mean-square error and standard error of pooled estimates, and for non-ignorable missing data it increases their bias.

For these reasons, I recommend that meta-analysts should not generally ignore within-study correlations, unless they truly are 0 or close to zero. This contradicts Ishak *et al.* (2008) who say that if interest lies only in the pooled effects then we can 'assume independence without any significant risk of bias or loss of precision in estimates'. Their recommendation is plausible only when

(a) the between-study variation is very large relative to the within-study variation, as then the between-study correlation, rather than the within-study correlation, is most influential (Section 3.1), or
(b) there are complete data with only small differences in the within-study covariance matrices across studies, as in this situation there is little opportunity for the pooled estimates to borrow strength anyway (Section 3.2).

Even in these situations, if one is interested in some function of the pooled estimates or their correlation, then within-study correlation may still be influential and should not be ignored. (Riley, Abrams, Lambert, Sutton and Thompson, 2007; Ishak *et al.*, 2008). For example, Higgins *et al.* (2008) recommend that mean effect estimates from a random-effect meta-analysis should be presented with a prediction interval, to reflect the model assumption that underlying study effects are drawn from some population of effects. This can be achieved in the multivariate setting by calculating a joint prediction interval for the end point effects in an individual study, which should account for the correlation between end points to be appropriate (Daniels and Hughes, 1997; Van Houwelingen *et al.*, 2002; Reitsma *et al.*, 2005).

Given that within-study correlation is important, it is especially unfortunate that published studies rarely report it. It is not, though, an insurmountable problem. Section 6 provides five options to proceed with multivariate meta-analysis when the within-study correlations are unknown, with one specific to multiple survival time points. Perhaps the best general approach to take is to obtain IPD and to estimate the within-study correlations directly; however, this may be time consuming and require complex statistical modelling, and IPD may not always be available. The most popular approach is a sensitivity analysis over an (informed) range of imputed correlations; for simplicity this usually assumes a common within-study correlation

across studies, but this may not always be realistic. Recently an alternative model for BRMA has been proposed which does not require the within-study correlations, as it rather models the overall correlation in the data (Riley, Thompson and Abrams, 2008). This model is appropriate when the pooled estimates (or some function of them), their precision or their correlation are of interest. Section 6.5 showed that it produces pooled estimates with better statistical properties than those from a general BRMA model ignoring within-study correlation. Extension of the alternative model for three or more end points is potentially important. I have focused on two outcomes in this paper, but the recommendation to account for within-study correlation extends to situations involving three or more outcomes, where then multiple within-study correlations per study are required.

Further research should also consider methods to assess publication bias and missingness assumptions within the multivariate setting. This issue has received little attention (Riley *et al.*, 2004; Jackson *et al.*, 2005) and multivariate meta-analysis references rather assess publication bias within the univariate framework (Nam *et al.*, 2003; Rothstein *et al.*, 2005). Univariate meta-analysis assumes that unknown studies are missing completely at random. Multivariate meta-analysis makes the additional assumption that, in known studies providing only a partial set of end points, the missing end point estimates are missing at random, i.e. that the observed relationship in those studies providing all end points is the same in those studies providing a partial set of end points. There are thus potentially two sources of missing data in a multivariate meta-analysis, and methods to assess their missingness assumptions need further development.

7.1. Link to the utilization of correlation in other areas

The importance of utilizing correlation is also apparent in other types of statistical models. For example, in econometrics the joint, rather than the separate, analysis of a number of correlated linear regressions is termed 'seemingly unrelated regression' (Zellner, 1962), and this approach produces more efficient parameter estimates (Zellner, 1962; Mehta and Swamy, 1976). In the joint synthesis of multiple treatment comparisons (Gleser and Olkin, 1994; Hasselblad, 1998; Caldwell *et al.*, 2005; Lu and Ades, 2004), correlation enables the utilization of direct and indirect evidence regarding a treatment benefit (Higgins and Whitehead, 1996; Lu and Ades, 2004; Salanti *et al.*, 2008), which can lead to more precise conclusions (Caldwell *et al.*, 2005). In longitudinal data models, it is necessary to account for the correlation between multiple patient responses, as otherwise the standard error of parameter estimates will be inappropriate (Dunlop, 1994). Furthermore, accounting for the correlation (latent association) between longitudinal data and survival data can reduce bias in parameter estimates (Henderson *et al.*, 2000), such as in the joint analysis of surrogate biomarkers and survival (Daniels and Hughes, 1997; Henderson *et al.*, 2002), which is particularly important given informative dropout (Thiebaut *et al.*, 2005).

7.2. Conclusion

Multivariate meta-analysis should play an important role in evidence-based decision making, as it appropriately accounts for the correlation between multiple end points and offers numerous advantages over the univariate approach. This paper has shown how and why within-study correlation influences pooled estimates from a BRMA, and that simply ignoring it can lead to meta-analysis results with inferior statistical properties. Options for dealing with unknown within-study correlations have also been reviewed, to help those facing this problem. Meta-analysis of cluster randomized trials (Donner and Klar, 2002), of crossover trials (Elbourne *et al.*, 2002) and of changes from baseline (Abrams *et al.*, 2005) have a similar problem of

unavailable correlations, yet it is still acknowledged that correlation should be accounted for in such syntheses. This message has gone somewhat awry within the context of meta-analysing multiple correlated end points, despite being first highlighted over 20 years ago (Raudenbush *et al.*, 1988). I thus strongly echo the above recommendation of Becker (2000) that meta-analysis must start to recognize and account for the correlation between multiple end points, as it is the most appropriate approach to take.

Acknowledgements

I thank the Associate Editor and a referee for making valuable comments that have helped considerably to improve the paper.

References

Abrams, K. R., Gillies, C. L. and Lambert, P. C. (2005) Meta-analysis of heterogeneously reported trials assessing change from baseline. *Statist. Med.*, **24**, 3823–3844.

Ades, A. E., Lu, G. and Higgins, J. P. (2005) The interpretation of random-effects meta-analysis in decision models. *Med. Decsn Makng*, **25**, 646–654.

Arends, L. R., Hunink, M. G. and Stijnen, T. (2008) Meta-analysis of summary survival curve data. *Statist. Med.*, **27**, 4381–4396.

Arends, L. R., Voko, Z. and Stijnen, T. (2003) Combining multiple outcome measures in a meta-analysis: an application. *Statist. Med.*, **22**, 1335–1353.

Becker, B. J. (2000) Multivariate meta-analysis. In *Handbook of Applied Multivariate Statistics and Mathematical Modeling* (eds H. E. A. Tinsley and S. Brown). San Diego: Academic Press.

Becker, B. J., Hedges, L. V. and Pigott, T. D. (2004) Campbell Collaboration statistical analysis policy brief. *Resource Document*. Campbell Collaboration. (Available from `http://www.campbellcollaboration.org/ECG/policy_stat.asp`.)

Berkey, C. S., Anderson, J. J. and Hoaglin, D. C. (1996) Multiple-outcome meta-analysis of clinical trials. *Statist. Med.*, **15**, 537–557.

Berkey, C. S., Antczak-Bouckoms, A., Hoaglin, D. C., Mosteller, F. and Pihlstrom, B. L. (1995) Multiple-outcomes meta-analysis of treatments for periodontal disease. *J. Dent. Res.*, **74**, 1030–1039.

Berkey, C. S., Hoaglin, D. C., Antczak-Bouckoms, A., Mosteller, F. and Colditz, G. A. (1998) Meta-analysis of multiple outcomes by regression with random effects. *Statist. Med.*, **17**, 2537–2550.

Berrington, A. and Cox, D. R. (2003) Generalized least squares for the synthesis of correlated information. *Biostatistics*, **4**, 423–431.

Browne, W. J. and Draper, D. (2000) Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Computnl Statist.*, **15**, 391–420.

Caldwell, D. M., Ades, A. E. and Higgins, J. P. (2005) Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *Br. Med. J.*, **331**, 897–900.

Chu, H. and Cole, S. R. (2006) Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J. Clin. Epidem.*, **59**, 1331–1333.

Daniels, M. J. and Hughes, M. D. (1997) Meta-analysis for the evaluation of potential surrogate markers. *Statist. Med.*, **16**, 1965–1982.

Dear, K. B. (1994) Iterative generalized least squares for meta-analysis of survival data at multiple times. *Biometrics*, **50**, 989–1002.

Donner, A. and Klar, N. (2002) Issues in the meta-analysis of cluster randomized trials. *Statist. Med.*, **21**, 2971–2980.

Dunlop, D. D. (1994) Regression for longitudinal data: a bridge from least squares. *Am. Statistn*, **48**, 299–303.

Elbourne, D. R., Altman, D. G., Higgins, J. P., Curtin, F., Worthington, H. V. and Vail, A. (2002) Meta-analyses involving cross-over trials: methodological issues. *Int. J. Epidem.*, **31**, 140–149.

Gail, M. H., Pfeiffer, R., Van Houwelingen, H. C. and Carroll, R. J. (2000) On meta-analytic assessment of surrogate outcomes. *Biostatistics*, **1**, 231–246.

Gelman, A. (2006) Prior distibutions for variance parameters in hierarchical models. *Bayes. Anal.*, **1**, 515–533.

Gentle, J. E. (1998) Cholesky factorization. In *Numerical Linear Algebra for Applications in Statistics* (ed. J. E. Gentle). Berlin: Springer.

Gleser, L. J. and Olkin, I. (1994) Stochastically dependent effect sizes. In *The Handbook of Research Synthesis* (eds H. Cooper and L. V. Hedges). New York: Russell Sage Foundation.

Harbord, R. M., Deeks, J. J., Egger, M., Whiting, P. and Sterne, J. A. (2007) A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics*, **8**, 239–251.

Hasselblad, V. (1998) Meta-analysis of multitreatment studies. *Med. Decisn Makng*, **18**, 37–43.
Hedges, L. V. and Olkin, I. (1994) *Statistical Methods for Meta-analysis*. Orlando: Academic Press.
Henderson, R., Diggle, P. and Dobson, A. (2000) Joint modelling of longitudinal measurements and event time data. *Biostatistics*, **1**, 465–480.
Henderson, R., Diggle, P. and Dobson, A. (2002) Identification and efficacy of longitudinal markers for survival. *Biostatistics*, **3**, 33–50.
Higgins, J. P., Thompson, S. G., Deeks, J. J. and Altman, D. G. (2003) Measuring inconsistency in meta-analyses. *Br. Med. J.*, **327**, 557–560.
Higgins, J. P., Thompson, S. G. and Spiegelhalter, D. J. (2009) A re-evaluation of random-effects meta-analysis. *J. R. Statist. Soc.* A, **172**, 137–159.
Higgins, J. P. and Whitehead, A. (1996) Borrowing strength from external trials in a meta-analysis. *Statist. Med.*, **15**, 2733–2749.
Ishak, K. J., Platt, R. W., Joseph, L. and Hanley, J. A. (2008) Impact of approximating or ignoring within-study covariances in multivariate meta-analyses. *Statist. Med.*, **27**, 670–686.
Ishak, K. J., Platt, R. W., Joseph, L., Hanley, J. A. and Caro, J. J. (2007) Meta-analysis of longitudinal studies. *Clin. Trials*, **4**, 525–539.
Jackson, D., Copas, J. and Sutton, A. J. (2005) Modelling reporting bias: the operative mortality rate for ruptured abdominal aortic aneurysm repair. *J. R. Statist. Soc.* A, **168**, 737–752.
Kalaian, H. A. and Raudenbush, S. W. (1996) A multivariate mixed linear model for meta-analysis. *Psychol. Meth.*, **1**, 227–235.
Komuro, H., Valentine, M. B., Rowe, S. T., Kidd, V. J., Makino, S., Brodeur, G. M., Cohn, S. L. and Look, A. T. (1998) Fluorescence in situ hybridization analysis of chromosome 1p36 deletions in human MYCN amplified neuroblastoma. *J. Pediatr. Surg.*, **33**, 1695–1698.
Korn, E. L., Albert, P. S. and McShane, L. M. (2005) Assessing surrogates as trial endpoints using mixed models. *Statist. Med.*, **24**, 163–182.
Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R. and Jones, D. R. (2004) How Vague is Vague?: a simulation study of the impact of the use of vague prior distributions in MCMC. *Statist. Med.*, **23**, 2401–2428.
Lu, G. and Ades, A. E. (2004) Combination of direct and indirect evidence in mixed treatment comparisons. *Statist. Med.*, **23**, 3105–3124.
Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000) WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statist. Comput.*, **10**, 325–337.
Mc Daid, C., Griffin, S., Weatherley, H., Durée, K., van der Burgt, M., van Hout, S., Akers, J., Davies, R. J. O., Sculpher, M. and Westwood, M. (2007) Sleep apnoea continuous positive airways pressure (CPAP) ACD: assessment report. *Report*. National Institute for Clinical Excellence, London. (Available from `http://guidance.nice.org.uk/page.aspx?o=280765`.)
Mehta, J. S. and Swamy, P. A. V. B. (1976) Further evidence on the relative efficiences of Zellner's seemingly unrelated regressions estimator. *J. Am. Statist. Ass.*, **71**, 634–639.
Nam, I. S., Mengersen, K. and Garthwaite, P. (2003) Multivariate meta-analysis. *Statist. Med.*, **22**, 2309–2333.
Raudenbush, S. W., Becker, B. J. and Kalaian, H. (1988) Modeling multivariate effect sizes. *Psychol. Bull.*, **103**, 111–120.
Reitsma, J. B., Glas, A. S., Rutjes, A. W., Scholten, R. J., Bossuyt, P. M. and Zwinderman, A. H. (2005) Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J. Clin. Epidem.*, **58**, 982–990.
Riley, R. D. (2005) Evidence synthesis of prognostic marker studies. *PhD Thesis*. University of Leicester, Leicester.
Riley, R. D., Abrams, K. R., Lambert, P. C., Sutton, A. J. and Thompson, J. R. (2007) An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Statist. Med.*, **26**, 78–97.
Riley, R. D., Abrams, K. R., Sutton, A. J., Lambert, P. C. and Thompson, J. R. (2007) Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Med. Res. Methodol.*, **7**, no. 3.
Riley, R. D., Burchill, S. A., Abrams, K. R., Heney, D., Lambert, P. C., Jones, D. R., Sutton, A. J., Young, B., Wailoo, A. J. and Lewis, I. J. (2003) A systematic review and evaluation of the use of tumour markers in paediatric oncology: Ewing's sarcoma and neuroblastoma. *Hlth Technol. Assessmnt*, **7**, no. 5.
Riley, R. D., Lambert, P. C., Staessen, J. A., Wang, J., Gueyffier, F., Thijs, L. and Boutitie, F. (2008) Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Statist. Med.*, **27**, 1870–1893.
Riley, R. D., Simmonds, M. C. and Look, M. P. (2007) Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. *J. Clin. Epidem.*, **60**, 431–439.
Riley, R. D., Sutton, A. J., Abrams, K. R. and Lambert, P. C. (2004) Sensitivity analyses allowed more appropriate and reliable meta-analysis conclusions for multiple outcomes when missing data was present. *J. Clin. Epidem.*, **57**, 911–924.
Riley, R. D., Thompson, J. R. and Abrams, K. R. (2008) An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics*, **9**, 172–186.
Rooney, B. L. and Murray, D. M. (1996) A meta-analysis of smoking prevention programs after adjustment for errors in the unit of analysis. *Hlth Educ. Q.*, **23**, 48–64.

Rothstein, H. R., Sutton, A. J. and Borenstein, M. E. (2005) *Publication Bias in Meta-analysis*. Chichester: Wiley.
Rubin, D. B. (1976) Inference and missing data. *Biometrika*, **63**, 581–592.
Salanti, G., Higgins, J., Ades, A. E. and Ioannidis, J. P. (2008) Evaluation of networks of randomized trials. *Statist. Meth. Med. Res.*, **17**, 279–301.
Sohn, S. Y. (2000) Multivariate meta-analysis with potentially correlated marketing study results. *Nav. Res. Logist.*, **47**, 500–510.
Spiegelhalter, D. J., Abrams, K. R. and Myles, J. P. (2003) *Bayesian Approaches to Clinical Trials and Health-care Evaluation*. Chichester: Wiley.
Spiegelhalter, D. J., Thomas, A. and Best, N. G. (2000) *WinBUGS Version 1.3 User Manual*. Cambridge: Medical Research Council Biostatistics Unit.
StataCorp (2007) *Statistical Software: Release 10.0*. College Station: Stata Corporation.
Stewart, L. A. and Tierney, J. F. (2002) To IPD or not to IPD?: advantages and disadvantages of systematic reviews using individual patient data. *Evaln Hlth Professnl*, **25**, 76–97.
Thiebaut, R., Jacqmin-Gadda, H., Babiker, A. and Commenges, D. (2005) Joint modelling of bivariate longitudinal data with informative dropout and left-censoring, with application to the evolution of CD4+ cell count and HIV RNA viral load in response to treatment of HIV infection. *Statist. Med.*, **24**, 65–82.
Thompson, J. R., Minelli, C., Abrams, K. R., Tobin, M. D. and Riley, R. D. (2005) Meta-analysis of genetic studies using Mendelian randomization—a multivariate approach. *Statist. Med.*, **24**, 2241–2254.
Van Houwelingen, H. C., Arends, L. R. and Stijnen, T. (2002) Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statist. Med.*, **21**, 589–624.
Van Houwelingen, H. C., Zwinderman, K. H. and Stijnen, T. (1993) A bivariate approach to meta-analysis. *Statist. Med.*, **12**, 2273–2284.
Zellner, A. (1962) An efficient method for estimating seemingly unrelated regressions and tests of aggregation bias. *J. Am. Statist. Ass.*, **57**, 500–509.