Research Article

Statistics in Medicine

Received 28 November 2011, Accepted 2 January 2013 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5745

Bayesian multivariate meta-analysis with multiple outcomes

Yinghui Wei*† and Julian P. T. Higgins

There has been a recent growth in developments of multivariate meta-analysis. We extend the methodology of Bayesian multivariate meta-analysis to the situation when there are more than two outcomes of interest, which is underexplored in the current literature. Our objective is to meta-analyse summary data from multiple outcomes simultaneously, accounting for potential dependencies among the data. One common issue is that studies do not all report all of the outcomes of interests, and we take an approach relying on marginal modelling of only the reported data. We employ a separation prior for the between-study variance–covariance matrix, which offers an improvement on the conventional inverse-Wishart prior, showing robustness in estimation and flexibility in incorporating prior information. Particular challenges arise when the number of outcomes is large relative to the number of studies because the number of parameters in the variance–covariance matrix can become substantial and there can be very little information with which to estimate between-study correlation coefficients. We explore assumptions that reduce the number of parameters in this matrix, including assumptions of homogenous variances, homogenous correlations for certain outcomes and positive correlation coefficients. We illustrate the methods with an example data set from the *Cochrane Database of Systematic Reviews*.

Keywords: Bayesian; multivariate meta-analysis; multiple outcomes; unconstrained parameterization; structured covariance matrix; missing outcome data

1. Introduction

Multivariate meta-analysis has been widely studied during the past two decades [1–13]. A multivariate approach is motivated by concerns of dependencies among treatment effect estimates and provides a natural refinement over a univariate approach, in which meta-analysis of each outcome separately is found to be practically inefficient and statistically imprecise [12, 14]. For a recent overview of advantages and limitations and for practical considerations, we refer the reader to Jackson *et al.* [15] and Mavridis and Salanti [16], respectively.

There has been a recent growth in developments of both methods [5, 7, 8, 10, 17] and statistical packages [18–21] for multivariate meta-analysis. A Bayesian approach is flexible and can readily be implemented using Markov chain Monte Carlo methods and powerful software such as WinBUGS [22] but has been investigated less extensively than frequentist methods.

In this paper, we consider Bayesian multivariate meta-analysis for the synthesis of summary data on multiple outcomes. We consider two important technical issues in the multivariate setting. First, because each study does not necessarily report all outcomes of interest, the data set is often incomplete. Second, the between-study variance and covariance terms are often difficult to estimate precisely. We address these issues particularly in the context of implementing multivariate meta-analyses when there are more than two outcomes.

Broadly speaking, there are two approaches to solving the incomplete data (missing outcomes) problem. In one approach, a data augmentation method is employed, in which a complete data set is created by inserting very small pieces of information for missing outcomes [23]. The variances for the imputed data are designed to be sufficiently large that the augmentation has no (or a trivial) influence

MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge, CB2 0SR, U.K.
**Correspondence to: Yinghui Wei, MRC Clinical Trials Unit, Aviation House, 125 Kingsway, WC2B 6NH, U.K. (present address).*
†*E-mail: y.wei@ctu.mrc.ac.uk*

in the analysis. The alternative approach, previously described for the situation of two outcomes [8] and which we pursue in this paper, uses a marginal likelihood approach based on the reported data. Specifically, the likelihood is a product of marginal distributions over all reported outcomes. Thus, for a meta-analysis with two outcomes, the likelihood is a mixture of univariate and bivariate normal distributions. This approach overcomes the problem that augmentation can potentially influence the analysis if not carefully checked. However, the model has only been applied to meta-analyses of two outcomes, and the difficulties of extension to more than two have been acknowledged [8], partly because of the constraint that the between-study variance–covariance matrix must be positive semidefinite.

Estimation difficulties for between-study variance have been noted [10, 15]. Here, we take an approach that models the variance–covariance matrix in terms of variances and correlations separately, employing unconstrained reparameterization of the variance–covariance matrix [24, 25]. This allows for robust estimation as well as the flexibility of incorporating prior information, for example using empirical evidence [26, 27]. Estimation of variances and covariances can be particularly difficult when the number of studies, N, is relatively small compared with the number of outcomes, p. Although information can be obtained by incorporating more outcomes into a model, through exploiting dependencies between them, this comes at the price of an increasing number of parameters being introduced into the between-study variance–covariance matrix. We consider models that reduce the number of parameters in the variance–covariance matrix by imposing structural assumptions, which impose (weakly) informative prior constraints on the variance–covariance matrix.

In the following section, we describe an example data set from the *Cochrane Database of Systematic Reviews* that we will use to illustrate the methods in the paper. Section 3 presents the underlying model used for multiple outcomes with incomplete data. Section 4 describes the Bayesian approach and prior distributions for fully flexible variance–covariance matrices, with application to the example data set provided in Section 5. Section 6 describes approaches for reducing the numbers of parameters in the variance–covariance matrix, considering a variety of assumptions that might be commonly applicable in practice, with application illustrated in Section 7.

2. Application: acute stroke data

Geeganage *et al.* [28] assessed the effect of vasoactive drugs for acute stroke in a systematic review. The review included 43 randomized controlled trials (7649 patients) of vasoactive drugs for acute ischaemic stroke or haemorrhagic stroke within 1 week of onset. Several different outcomes were collected, including death by the end of the trial (D), death or deterioration by the end of the trial (DD), systolic blood pressure (SBP) after 24- to 72-h treatment and diastolic blood pressure (DBP) at the same time point. Some trials had more than one vasoactive drug group, and the total number of comparisons in our analysis is 47 (the review authors divided the control group among multiple comparisons when there was more than one treatment group). For simplicity, we will henceforward refer to these as 47 trials.

Table 1 shows whether outcomes are reported by specific trials. The number of trials reporting on each individual outcome varies from 29 to 41. Only 21 of the trials reported all four of the outcomes. Note that the outcome of death is nested within the compound outcome of death or disability, and SBP and DBP are highly correlated [29]. We will perform multivariate meta-analysis, by which we take into account the dependencies between treatment effect estimates for these outcomes.

Table I. Outcomes reported by trials of vasoactive drugs for acute stroke.

	Number of trials							
Outcomes reported	21	8	7	1	5	4	1	Total
Death	✓	✓	✓	✓	×	✓	×	41
Death or disability	✓	×	✓	×	×	×	✓	29
SBP	✓	✓	×	✓	✓	×	×	35
DBP	✓	✓	×	×	✓	×	×	34

Statistics in Medicine

3. Bayesian modelling of multiple outcomes

We consider the meta-analysis of N studies with p outcomes of interest. We write as p_i the number of outcomes examined by study i, $i = 1, 2, \ldots, N$. Studies do not necessarily report all outcomes of interest, indicating that $1 \leqslant p_i \leqslant p$. We refer to absence of data for one or more outcomes as study-level missing data.

We consider multivariate normal models with a mixture of different dimensions, p_i, $i = 1, 2, \ldots, N$. Specifically, for study i, we assume

$$\begin{aligned} \mathbf{y}_i &\sim \mathrm{MVN}\left(\boldsymbol{\theta}_i, \Sigma_{i_i}\right), \\ \boldsymbol{\theta}_i &\sim \mathrm{MVN}\left(\mathrm{X}_i\boldsymbol{\mu}, \mathrm{X}_i\Omega\mathrm{X}_i^{\mathrm{T}}\right). \end{aligned} \tag{1}$$

The two formulae respectively represent a multivariate normal likelihood for the observed data and a random-effects assumption, which is also a multivariate normal. Vector $\mathbf{y}_i$ is a $p_i \times 1$ ($p_i \geqslant 1$) vector of treatment effect estimates, $\boldsymbol{\theta}_i$ is a $p_i \times 1$ vector of random effects, Σ_i is a $p_i \times p_i$ matrix with within-study variances s_{ii}^2 as diagonal elements and within-study covariances as nondiagonal elements, X_i is a $p_i \times p$ design matrix defining which of the p outcomes are included in the study, $\boldsymbol{\mu}$ is a $p \times 1$ vector of underlying mean treatment effects across studies and Ω is a $p \times p$ matrix representing the between-study covariance matrix for all p outcomes of interest. For studies that include only one outcome, the distributions in (1) are univariate. We assume that the within-study variances and covariances are known and uncorrelated with the treatment effects. These are standard assumptions in meta-analysis, which, although unlikely to have substantial implications, should be assessed critically in practical applications. The design matrices X_i facilitate the modelling of studies with various missing data (or missing outcome) patterns. This technique is analogous to that proposed by Glester and Olkin in [30] and used by Lu *et al.* [31].

In our example data, we consider a meta-analysis with $p = 4$. For illustration of model (1), we suppose that study i reports outcomes 1, 3 and 4 and so contributes a vector of estimated treatment effects $\mathbf{y}_i = (y_{i1}, y_{i3}, y_{i4})$. Then, the model for study i is

$$\begin{pmatrix} y_{i1} \\ y_{i3} \\ y_{i4} \end{pmatrix} \sim \mathrm{MVN}\left(\begin{pmatrix} \theta_{i1} \\ \theta_{i3} \\ \theta_{i4} \end{pmatrix}, \begin{bmatrix} s_1^2 & r_{12}s_1s_2 & r_{13}s_1s_3 \\ . & s_2^2 & r_{23}s_2s_3 \\ . & . & s_3^2 \end{bmatrix}\right),$$

$$\begin{pmatrix} \theta_{i1} \\ \theta_{i3} \\ \theta_{i4} \end{pmatrix} \sim \mathrm{MVN}\left(\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\right.$$

$$\left. \times \begin{bmatrix} \tau_1^2 & \rho_{12}\tau_1\tau_2 & \rho_{13}\tau_1\tau_3 & \rho_{14}\tau_1\tau_4 \\ . & \tau_2^2 & \rho_{23}\tau_2\tau_3 & \rho_{24}\tau_2\tau_4 \\ . & . & \tau_3^2 & \rho_{34}\tau_3\tau_4 \\ . & . & . & \tau_4^2 \end{bmatrix}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^{\mathrm{T}}\right),$$

where r_{ij} and ρ_{ij} represent within-study and between-study correlation coefficients between outcomes ($i, j = 1, 2, 3; i < j$) and τ_j represents heterogeneity between studies for outcome j($j = 1, 2, 3, 4$).

The studies are linked through the parameters that characterize the distribution of the random effects, $\boldsymbol{\mu}$ and Ω. The estimation of parameters $\boldsymbol{\mu}$ and Ω is one of the objectives of a multivariate random-effects meta-analysis. The Bayesian approach requires the specification of prior distributions for unknown parameters $\boldsymbol{\mu}$ and Ω. For $\boldsymbol{\mu}$, a conjugate prior for treatment effects is a multivariate normal distribution, and we opt for an approximately noninformative distribution, with prior independence of the p treatment effects,

$$\boldsymbol{\mu} \sim \mathrm{MVN}(0, 1000I_p),$$

where I_p is the $p \times p$ identity matrix. The specification of the prior distribution for the between-study covariance matrix is more challenging and is the subject of much of the rest of the paper.

4. Unstructured variance–covariance matrix

The most flexible structure for the variance–covariance matrix is that which makes the fewest assumptions, and we refer to this as an unstructured variance–covariance matrix. The assumption of an unstructured matrix facilitates the use of a prior distribution from the conjugate family.

4.1. Conjugate priors

For the multivariate normal models in (1), the conjugate prior distribution for the between-study variance–covariance matrix is an inverse-Wishart distribution [32],

$$\Omega \sim \mathrm{IW}(V, k), \tag{2}$$

where V is a $p \times p$ scale matrix and k ($\geq p$) the number of degrees of freedom. The matrix V is constrained to be positive semidefinite. The scale matrix not only affects the location of the Wishart distribution but also, jointly with k, determines the dispersion of the distribution. Conventionally, the scale matrix is chosen to reflect a prior estimate of the variance–covariance matrix. The number of degrees of freedom is chosen to be as small as possible to reflect vague prior knowledge [33]. Note that in the Wishart distribution, the number of degrees of freedom, k, is constrained to be no less than p, the rank of the scale matrix, which here equals the number of outcomes. This implies that the higher the number of outcomes, the less vague is the Wishart prior, indicating a potential influence of the Wishart prior in the posterior estimates.

Figure 1 illustrates the influence of the magnitude of the diagonal elements of the scale matrix on posterior estimates of the between-study variance–covariance matrix, using the example data. The figure shows a sensitivity analysis using the model described in (1) using inverse-Wishart prior distributions for Ω, with degrees of freedom of $k = 4$, based on 1000 different diagonal scale matrices. Each prior scale

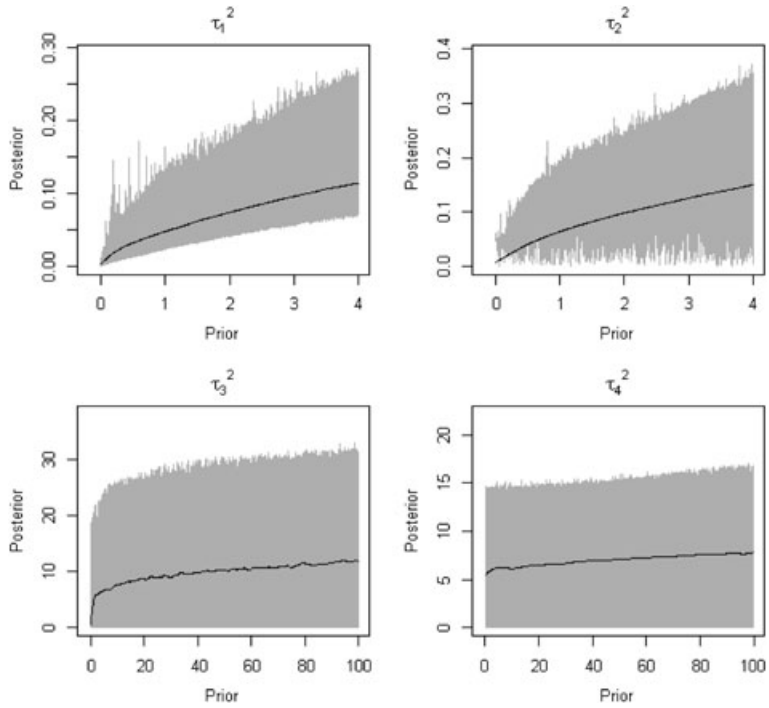


Figure 1. Sensitivity analysis for multivariate meta-analysis using Wishart prior distributions (degrees of freedom = 4) with 1000 different scale matrices, applied to the acute stroke data. The horizontal axis represents the prior value of the between-study variance and the vertical axis the posterior estimate of between-study variance estimated from the multivariate meta-analysis using the Wishart prior distribution corresponding to the prior value. The solid line is the posterior median, with shaded regions describing the associated 95% CI. The between-study variances for death (D), death or deterioration (DD), systolic blood pressure (SBP) and diastolic blood pressure (DBP) are denoted by τ_1^2, τ_2^2, τ_3^2 and τ_4^2, respectively.

matrix is obtained by sampling between-study standard deviations for each of the four outcomes. Specifically, we simulate τ_{j0} from uniform(0, 2) for $j = 1$ or 2 (dichotomous outcomes) or from uniform(0, 10) for $j = 3$ or 4 (continuous outcomes). These are the between-study standard deviations that we expect to obtain when randomly drawing a sample from a Wishart distribution with degrees of freedom set to 4. Denoting by v_j^2 the diagonal elements of the scale matrix V $(j = 1, \ldots, 4)$, we use a property of the Wishart distribution, that $E(\tau_j^2) = k v_j^2$, to calculate the values of v_j^2 using the sampled prior values $E(\tau_j^2) = \tau_{j0}^2$. Given the variance of the Wishart distribution, $\text{var}(\tau_j^2) = 2k v_j^4$, we note that the variance of the prior distribution is influenced by v_j if k is fixed. Because $E(\tau_j^2) = k v_j^2$, the smaller the prior guess of v_j, the more informative is the prior distribution on τ_j.

We see in the figure that, in general, the estimates of between-study variance are sensitive to the scale matrix. The sensitivity is most obvious for the two dichotomous outcomes, where the posterior median for the heterogeneity parameters increase from 0 and continues to increase as the prior mean increases. The sensitivity analysis of the example shows that the Wishart prior distribution is very influential in the estimation of the between-study variance–covariance matrix. The Wishart distribution generalizes the gamma distribution, and our finding is similar to what has been observed in the univariate setting where a gamma prior distribution leads to overestimation of the heterogeneity parameter when the true value is close to 0 [34]. Wishart priors can become dominant in the posterior distribution as the dimension of the matrix becomes higher because there are more parameters to estimate and the Wishart prior becomes more informative when p becomes larger.

The influential impact of the prior distribution on the posterior distribution is a key limitation of the Wishart prior in multivariate meta-analysis. Furthermore, explicitly representing an informative prior distribution is difficult. The disadvantages of the inverse-Wishart prior distribution have been discussed and alternative prior distributions proposed [35–41]. Variance–covariance matrix estimation is important in hierarchical models because the distribution of random effects depends on estimation of the variance–covariance matrix. We therefore explore alternative strategies for specifying prior distributions for this matrix in the following sections.

4.2. *Separation strategies*

An alternative to a Wishart prior is to model the variance–covariance matrix in terms of standard deviations and correlations separately. The between-study variance–covariance matrix Ω can be written as

$$\Omega = V^{1/2} R V^{1/2},$$

where $V^{1/2}$ is a diagonal matrix with standard deviations as elements and R is the $k \times k$ matrix of correlations. This decomposition is referred to as a separation strategy by Barnard *et al.* [35]. Separating out the standard deviations and correlations allows independent prior distributions to be placed on them. We further reparameterize the correlation matrix to enforce the positive semidefinite constraints. We describe two different parameterizations that can be used within this separation strategy.

4.2.1. Separation strategy by Cholesky decomposition. Because the correlation matrix R is symmetric and positive semidefinite, it can be factored as

$$R = L^{\mathrm{T}} L, \tag{3}$$

where L is a $p \times p$ upper-triangular matrix. We set L_{ij} $(i = 1, \ldots, p,\ i \leqslant j)$ to denote the $p(p+1)/2$ elements in matrix L, referred to as the Cholesky factors.

Because R is a correlation matrix, its elements must lie in the range $[-1, 1]$. This induces constraints on possible values for the Cholesky factors L_{ij}, as explained in more detail in Appendix A. The diagonal entries of R are each 1, so $\rho_{kk} = \mathbf{L}_{k\cdot}^{\mathrm{T}} \mathbf{L}_{k\cdot} = 1$, where $\mathbf{L}_{k\cdot} = (L_{k1}, L_{k2}, \ldots, L_{kk})^{\mathrm{T}}$ denotes column k in matrix L. The Cholesky factors must lie in the intersection of $[-1, 1]$ and $\left[L_{ij}^l, L_{ij}^r\right]$, where $\left[L_{ij}^l, L_{ij}^r\right]$ satisfies the conditions that $\rho_{kk} = \mathbf{L}_{k\cdot}^{\mathrm{T}} \mathbf{L}_{k\cdot} = 1$. The prior distributions are then placed on the Cholesky factors L_{ij}.

In our example data, the correlation matrix has a dimension of 4, so R can be written as

$$R = \begin{bmatrix} 1 & L_{11}L_{12} & L_{11}L_{13} & L_{11}L_{14} \\ \cdot & L_{12}^2 + L_{22}^2 & L_{12}L_{13} + L_{22}L_{23} & L_{12}L_{14} + L_{22}L_{24} \\ \cdot & \cdot & L_{13}^2 + L_{23}^2 + L_{33}^2 & L_{13}L_{14} + L_{23}L_{24} + L_{33}L_{34} \\ \cdot & \cdot & \cdot & L_{14}^2 + L_{24}^2 + L_{34}^2 + L_{44}^2 \end{bmatrix}. \tag{4}$$

To ensure uniqueness, we set L_{11} to be 1. We show in Appendix A.1 that the plausible intervals for the remaining Cholesky factors are as follows:

$$\begin{aligned} &L_{12}, L_{13}, L_{14} \in [-1, 1], \\ &L_{22} = \sqrt{1 - L_{12}^2}, \\ &L_{23} \in \left[-\sqrt{1 - L_{13}^2}, \sqrt{1 - L_{13}^2}\right], \\ &L_{24} \in \left[-\sqrt{1 - L_{14}^2}, \sqrt{1 - L_{14}^2}\right], \\ &L_{33} = \sqrt{1 - L_{13}^2 - L_{23}^2}, \\ &L_{34} \in \left[-\sqrt{1 - L_{14}^2 - L_{24}^2}, \sqrt{1 - L_{14}^2 - L_{24}^2}\right], \\ &L_{44} = \sqrt{1 - L_{14}^2 - L_{24}^2 - L_{34}^2}. \end{aligned} \tag{5}$$

Note that the intervals in (5) were derived under the condition that $\rho_{kk} = \mathbf{L}_{k\cdot}^{\mathrm{T}}\mathbf{L}_{k\cdot} = 1$, ensuring that the positive semidefinite constraint for the correlation matrix is satisfied. To specify the prior distribution for the L_{ij} values, we place uniform distributions on each of these ranges. The use of these prior distributions for L_{ij} in (5) results in the preceding correlation matrix R with correlations $\rho_{ij} \in (-1, 1)$. The shape of the prior distributions for the individual correction coefficients is illustrated in Figure 2.

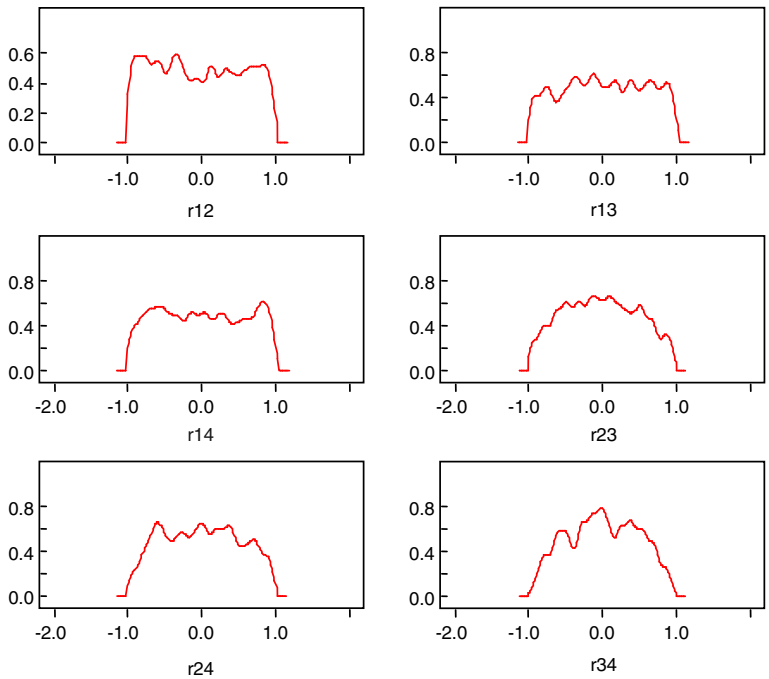


Figure 2. Prior distributions for between-study correlation from Cholesky parameterization.

The generalization of the prior distribution for unconstrained estimation of variance–covariance matrix for an arbitrary number, p, of outcomes is provided by

$$\begin{aligned} L_{1j} &\sim \text{uniform}(-1, 1), \\ L_{jj} &= \sqrt{1 - \sum_{i=1}^{j-1} L_{ij}^2}, \\ L_{ij} &\sim \text{uniform}\left(-\sqrt{1 - \sum_{k=1}^{i-1} L_{kj}^2}, \sqrt{1 - \sum_{k=1}^{i-1} L_{kj}^2} \right), \; i < j, \end{aligned} \tag{6}$$

for $j = 2, \ldots, p$, with $L_{11} = 1$. In these prior distributions, we place uniform distributions on the permissible intervals defined in (5), which ensures that the prior probability is 0 for the L_{ij} parameters taking values outside the intervals. The posterior probability is therefore 0 for the positive semidefinite constraint being violated.

4.2.2. Separation strategy by spherical decomposition. Lu and Ades [42] introduced spherical decomposition of the variance–covariance matrix into the field of meta-analysis. They applied the technique to mixed-treatment comparisons (or network meta-analysis), where correlation arises when treatment effects share the same baseline treatment. Here, we employ the decomposition technique in multiple-outcome meta-analysis. Spherical decomposition is a reparameterization of the Cholesky decomposition by using sine and cosine functions for L_{ij}. The products of these sine and cosine functions lie within the interval $[-1, 1]$, making them inherently convenient for parameterization of a correlation matrix. The reparameterization is again derived so that the products of diagonal entries satisfy the condition $\rho_{kk} = \mathbf{L}_{k.}^{\mathrm{T}} \mathbf{L}_{k.} = 1$, where $\mathbf{L}_{k.}$ denotes column k in (now spherical decomposition) matrix L. More specifically, we set $L_{11} = 1$ and for $k = 2, \ldots, p$,

$$\begin{aligned} L_{k1} &= \cos(\phi_{k2}) \\ L_{k2} &= \sin(\phi_{k2}) \cos(\phi_{k3}) \\ &\vdots \\ L_{k,k-1} &= \sin(\phi_{k2}) \sin(\phi_{k3}) \cdots \cos(\phi_{kk}) \\ L_{k,k} &= \sin(\phi_{k2}) \sin(\phi_{k3}) \cdots \sin(\phi_{kk}). \end{aligned} \tag{7}$$

To ensure the uniqueness of the spherical parameterization, we must have $\phi_{km} \in (0, \pi)$, for $m = 2, \ldots, k$ [25]. Then, the (i, j) entry of R can be represented as the inner product $\rho_{ij} = \mathbf{L}_{i.}^{\mathrm{T}} \mathbf{L}_{j.}$.

For illustration, the 4×4 upper-triangular matrix L for spherical decomposition is given by

$$L = \begin{pmatrix} 1 & \cos(\varphi_{21}) & \cos(\varphi_{31}) & \cos(\varphi_{41}) \\ 0 & \sin(\varphi_{21}) & \sin(\varphi_{31}) \cos(\varphi_{32}) & \sin(\varphi_{41}) \cos(\varphi_{42}) \\ 0 & 0 & \sin(\varphi_{31}) \sin(\varphi_{32}) & \sin(\varphi_{41}) \sin(\varphi_{42}) \cos(\varphi_{43}) \\ 0 & 0 & 0 & \sin(\varphi_{41}) \sin(\varphi_{42}) \sin(\varphi_{43}) \end{pmatrix}.$$

We assign noninformative prior distributions to parameters in our model as $\phi_{km} \sim \text{uniform}(0, \pi)$, which results in an unstructured correlation matrix R with $\rho_{ij} \in (-1, 1)$. Figure 3 illustrates the shape of the prior distributions for the individual correction coefficients.

5. Application (1): unstructured variance–covariance matrix

We applied multivariate models on the basis of (1) to the acute stroke data for illustration, using prior distributions described in Sections 4.1 and 4.2. Within-study covariances were imputed using the methods of Wei and Higgins [43], who derived expressions for covariances between treatment effect estimates as functions of correlations between the underlying outcome measures. Specifically, we used formula (10) and Equations (1.1) and (1.3) from Wei and Higgins' Table 1, with correlation coefficients between outcomes imputed as 0.74 between SBP and DBP (on the basis of external evidence) and 0.5 between pairs of blood pressure and mortality/morbidity outcomes. The correlation between treatment effect esti-

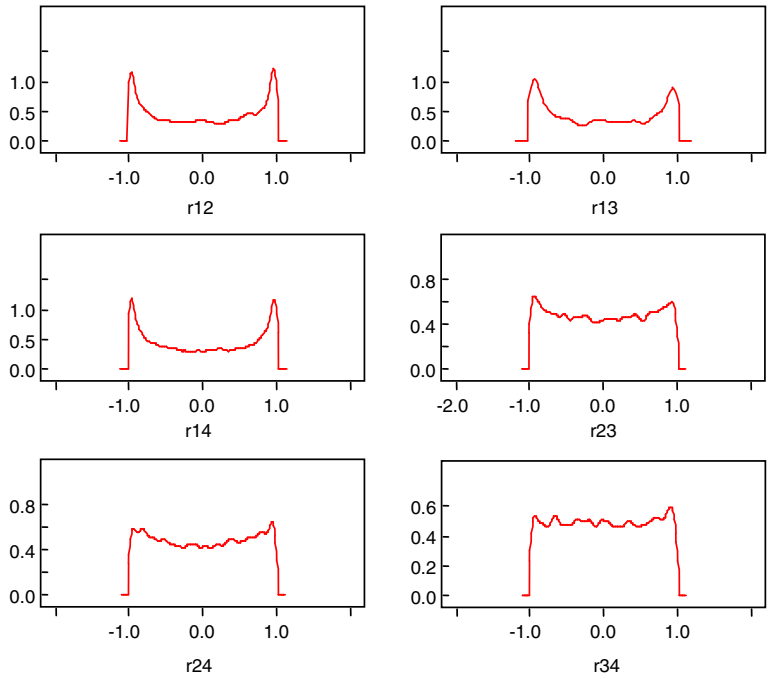


Figure 3. Prior distributions for between-study correlation from spherical parameterization.

mates for D and DD does not require any imputation because the former outcome is nested within the latter [43].

In the multivariate meta-analysis not using the separation strategy, we use the conjugate inverse-Wishart prior in (2) for the between-study variance–covariance matrix. In the univariate meta-analysis and the multivariate meta-analysis using the separation strategy, we use uniform(0,2) and uniform(0,10) as a prior distribution for the between-study standard deviation τ for dichotomous and continuous outcomes, respectively. This prior provides a sufficiently large range to cover reasonable values for the parameters and starts from 0 to cover the possibility that τ is close or equal to 0 (the latter corresponding to a fixed-effect meta-analysis model). The upper limit of 2 is based on predictive distributions for heterogeneity variances of log odds ratios, based on many meta-analyses, where 95% upper limits of around 2 were observed [27]. These uniform prior distributions will not violate the positive semidefinite constraint, which requires the standard deviation τ to be not less than 0.

We then implemented the models using MCMC techniques using WinBUGS [22], via the R2WinBUGS package [44]. The WinBUGS code is available in Appendix B. The convergence of MCMC algorithms was monitored by examining time series plots of the parameters over iterations and using the modified Gelman and Rubin approach for multiple chains [45]. For models with conjugate prior distributions, the convergence was quick, and the mixing was good. Analyses took longer to run if a separation prior was used, which is to be expected because elements of the variance–covariance matrix are split up and assigned independent priors, with the posterior distribution being updated one by one for each of the elements. Our analysis was based on 100 000 MCMC samples with the first 10 000 iteration discarded as burn-in and a thinning interval of 10 to reduce the effects of within-sample autocorrelation. Table II gives summary statistics from posterior distributions in multivariate meta-analyses with a Wishart prior distributions and using separation strategies with Cholesky and spherical parameterizations and includes comparisons with Bayesian univariate analyses. Plots of pairwise correlations between posterior estimates of effect sizes are available in Appendix C. [28]

The point estimates of treatment effect broadly agree between univariate and multivariate models. Both the Cholesky and spherical decompositions yield gains in precision in the multivariate approach compared with the univariate Bayesian approach, although the actual gains in precision are small. Variance terms are estimated to be similar between the two separation strategies. There is a substantial discrepancy between posterior medians from these strategies compared with using the Wishart prior distribution, reflecting the influence of the Wishart prior. The precision of estimation of between-study

Table II. Comparison of results of Bayesian multivariate meta-analyses with unstructured between-study variance–covariance matrix for acute stroke data.

	Univariate		Multivariate: inverse-Wishart		Multivariate: Cholesky		Multivariate: spherical	
Parameter of interest	Posterior median [95% CI]	Posterior SD	Posterior median [95% CI]	Posterior SD	Posterior median [95% CI]	Posterior SD	Posterior median [95% CI]	Posterior SD
μ_1	1.066 [0.91, 1.25]	0.08	1.00 [0.81, 1.21]	0.11	1.05 [0.92, 1.22]	0.08	1.05 [0.93, 1.22]	0.07
μ_2	1.11 [0.93, 1.33]	0.09	1.02 [−0.81, 1.28]	0.12	1.05 [0.90, 1.23]	0.08	1.05 [0.91, 1.24]	0.08
μ_3	−4.47 [−6.62, −2.41]	1.05	−4.157 [−5.84, −2.46]	0.87	−4.02 [−5.92, −1.99]	1.02	−3.87 [−5.77, −1.87]	1.02
μ_4	−2.57 [−3.97, −1.21]	0.70	−2.691 [−3.96, −1.41]	0.65	−2.53 [−3.84, −1.21]	0.67	−2.46 [−3.77, −1.11]	0.67
τ_1^2	0.011 [0.000, 0.11]	0.03	0.133 [0.065, 0.30]	0.06	0.004 [0.000, 0.075]	0.03	0.007 [0.000, 0.136]	0.03
τ_2^2	0.019 [0.000, 0.18]	0.05	0.16 [0.07, 0.37]	0.08	0.012 [0.000, 0.129]	0.04	0.017 [0.000, 0.19]	0.05
τ_3^2	14.22 [4.57, 38.35]	8.77	8.31 [2.37, 22.52]	5.36	13.37 [4.48, 33.69]	7.72	15.06 [5.51, 41.43]	9.47
τ_4^2	7.68 [3.04, 18.29]	3.93	6.933 [2.73, 15.6]	3.35	7.62 [3.25, 17.18]	3.59	8.78 [3.80, 20.17]	4.32
ρ_{12}	—	—	0.25 [−0.31, 0.68]	0.25	0.21 [−0.94, 0.98]	0.58	0.49 [−0.98, 1.00]	0.63
ρ_{13}	—	—	0.017 [−0.54, 0.57]	0.29	−0.12 [−0.82, 0.88]	0.63	−0.51 [−1.00, 1.00]	0.75
ρ_{14}	—	—	0.017 [−0.52, 0.57]	0.29	−0.10 [−0.97, 0.96]	0.62	−0.39 [−1.00, 1.00]	0.73
ρ_{23}	—	—	0.019 [−0.59, 0.58]	0.31	0.102 [−0.82, 0.88]	0.48	0.180 [−0.92, 0.94]	0.55
ρ_{24}	—	—	0.002 [−0.57, 0.57]	0.30	0.01 [−0.82, 0.86]	0.45	0.08 [−0.96, 0.95]	0.53
ρ_{34}	—	—	0.95 [0.75, 0.99]	0.07	0.79 [0.29, 0.96]	0.17	0.87 [0.45, 0.99]	0.14
$\bar{D}$	459.17		457.384		461.28		461.30	
pD	40.08		50.84		40.50		40.84	
DIC	535.26		508.22		501.78		502.15	
Run time (s)	51		64		27 076		15 559	

Bayesian results are based on 100 000 MCMC samples with first 10 000 iterations discarded as burn-in and thinning to every 10th sample. The deviance information criterion (DIC) is a statistics for Bayesian model comparison. It measures model fit by considering penalty for model complexity. Roughly, a difference of more than 10 might definitely rule out the model with the higher DIC. $\bar{D}$ is the posterior mean deviance; pD is the effective number of parameters.

Statistics in Medicine

variances is similar across approaches, although higher for the separation strategies. Between-study correlations are estimated very imprecisely, as is often the case in multivariate meta-analysis. However, the imprecise estimation of these correlations seems to have negligible impact on the treatment effect estimates, which are found to be similar across methods and priors used. The imprecise correlations would, in contrast, affect the prediction distribution for the true treatment effects in a new study, obtained as

$$\boldsymbol{\theta}_{\text{new}} \sim \text{MVN}\left(X_{\text{new}}\hat{\boldsymbol{\mu}}, X_{\text{new}}\hat{\Omega}X_{\text{new}}^{\text{T}}\right) \tag{8}$$

because the estimated variance–covariance matrix is used for this. We will explore in the next section whether estimation of variance–covariance matrix can be improved by making assumptions about the matrix structure.

We compared the models by computing the deviance information criterion (DIC) [46]. Multivariate models have substantially lower DIC values than the univariate model, indicating a preference for the multivariate model in terms of the balance between model fit and complexity. Multivariate models estimate the overall effects for multiple outcomes simultaneously and borrow strength not only across studies but also across outcomes. Some may argue that the multivariate methods are to be preferred in general; however, this would need to be balanced against the additional complexity of the analysis in practice. Analyses using the separation strategies have lower DICs than that using a Wishart prior. However, DIC values do not provide any evidence that one decomposition strategy is better than the other.

6. Structured variance–covariance matrices

An unstructured between-study variance–covariance matrix poses a substantial estimation problem when the number of outcomes is large. It introduces many parameters, and the amount of information about these, particularly the correlation coefficients, is likely to be minimal. Reducing model complexity in terms of variance–covariance structure is then appealing. Furthermore, if an outcome is addressed by only one study, then a random-effects meta-analysis model cannot be applied to it without either introducing some structure or imposing a strong prior distribution on the heterogeneity variance for that outcome. If there are only two or three studies, the heterogeneity parameters can be very imprecisely estimated. However, estimates can be improved to have tighter 95% intervals if we borrow strength from similar outcomes or use informative priors [27].

In this section, we consider a variety of structures for variance–covariance matrices in an attempt to obtain robust, yet plausible, estimates for the between-study variance–covariance matrix in the common situation of the number of studies, N, being small relative to the number of outcomes, p.

6.1. Structure I: homogeneous between-study variance for dichotomous outcomes

A common application of multivariate meta-analysis is to network meta-analysis, and a standard assumption in this situation is that the between-study variances are homogenous across comparisons [47, 48]. Such an assumption may be reasonable in a network meta-analysis situation because the same effect metric is used for all treatment effects (for example, a log odds ratio or a difference in means). In multiple-outcome meta-analysis, the assumption may be unreasonable, as is clearly the case in our example, which has a mixture of log odds ratios and differences in means for different continuous outcome measures. In particular, it is not realistic to assume that heterogeneity parameters are homogenous between dichotomous and continuous outcomes or for different measurement scales among continuous outcomes. For outcomes that do use the same treatment effect measure, however, we could assume that heterogeneity parameters are homogeneous across them. For instance, we might set $\tau_j = \tau_k = \tau_0$ if outcomes j and k are both dichotomous and analysed using log odds ratios and assign a single prior distribution of $\tau_0 \sim \text{uniform}(0, 2)$ for these. We implement such an assumption (referred to as structure I) in our example data by assuming this to be the case for death (log odds ratio θ_1) and death or deterioration (log odds ratio θ_2). However, we acknowledge that such an assumption is strong and is likely to be used only when outcomes are closely related or when there are many parameters and the model is at risk of being overfitted otherwise.

6.2. *Structure II: homogeneous between-study correlation for some outcomes*

Between-study correlation parameters contribute a significant number of parameters to the covariance matrix. Reducing the number of between-study correlation coefficients is therefore particularly desirable. Alternatively, or in addition to, assuming that some between-study variances are equal, we could assume that some between-study correlations are equal. This assumption may be more justifiable when pairs of outcomes are of similar types, for example, we might reasonably assume that the correlation between D and SBP is the same as the correlation between D and DBP. More specifically, in our example data, we assume a structured correlation matrix,

$$R = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \cdot & 1 & \rho_2 & \rho_3 \\ \cdot & \cdot & 1 & \rho_4 \\ \cdot & \cdot & \cdot & 1 \end{bmatrix}, \tag{9}$$

where we assume that the between-study correlation coefficient between D and SBP is identical to that between DD and SBP and we also assume that the between-study correlation coefficient between D and DBP is identical to that between DD and DBP. This structure reduces the number of correlation parameters from six to four.

We further obtained a Cholesky parameterization for this assumed structure for the correlation matrix. We demonstrate in Appendix A.2 how we derive the following priors:

$$\begin{aligned} L_{12} &\sim \text{uniform}(-1, 1), \\ L_{23}, L_{24} &\sim \text{uniform}\left(-\sqrt{\frac{1-L_{12}}{2}}, \sqrt{\frac{1-L_{12}}{2}}\right), \\ L_{34} &\sim \text{uniform}\left(-\sqrt{1-\frac{2}{(1-L_{12})}L_{24}^2}, \sqrt{1-\frac{2}{(1-L_{12})}L_{24}^2}\right). \end{aligned} \tag{10}$$

With the Cholesky parameterization for structured correlation (9) and the prior distributions in (10), the resulting between-study correlations are given by

$$\begin{aligned} \rho_1 &= L_{12}, \\ \rho_2 &= \frac{\sqrt{1-L_{12}^2}}{1-L_{12}} L_{23}, \\ \rho_3 &= \frac{\sqrt{1-L_{12}^2}}{1-L_{12}} L_{24}, \\ \rho_4 &= \frac{\sqrt{1-L_{12}^2}}{1-L_{12}} L_{12}L_{24} + +L_{23}L_{24} + L_{34}\sqrt{1-\rho_2^2-L_{23}^2}. \end{aligned}$$

The assumed structured correlation (9) can also, in principle, be decomposed by using spherical parameterization. However, this requires calculation of inverse trigonometric function for sines and cosines, which are not currently possible in the WINBUGS software. We therefore implement only a Cholesky parameterization for structure II.

6.3. *Structure III: positive between-study correlations for all outcomes*

In some circumstances, random effects within a trial can be assumed to be positively correlated without making the stronger assumption that they are equal. Specifically, in randomized trials, treatment effects for intended (beneficial) outcomes affected by the same mechanism of action (for example, the same biological pathway) are likely to be positively correlated. This suggests a narrowing of the prior distribution by restriction to only positive values, that is, to $\rho_{ij} \sim \text{uniform}(0, 1)$ rather than $\rho_{ij} \sim \text{uniform}(-1, 1)$. The structure allows correlations to vary freely across the narrower interval (0, 1). In the spherical decomposition framework, the equivalent restriction is to change the prior distribution for the φ parameters to be $\varphi_{ij} \sim \text{uniform}(0, \pi/2)$ [42]. In the Cholesky decomposition framework, the restriction is satisfied if we constrain all Cholesky factors L_{ij} to be greater than 0.

7. Application (2): structured variance–covariance matrix

Table III gives summary statistics from posterior distributions assuming various structures for the between-study variance–covariance matrix: structure I, structure II, structure III and the combinations of structures I and II, structures I and III and structures II and III. For treatment effect estimates, the posterior medians are similar across the models with different structures. Compared with the results from unstructured matrices in Table II, there is no big reduction in width of the 95% CIs, although some small gains in precision are evident in Table III.

However, assume that structured between-study variance–covariance matrices have an effect on the estimates of the elements of the matrix, with precision increasing for τ_1 (D) and τ_2 (DD) when a homogenous variance is assumed for these dichotomous outcomes (structure I) and precision increasing for correlations when some correlations are assumed to be identical (structure II) and when the prior distribution for correlation coefficients is restricted to the positive range (structure III).

Posterior medians and 95% CIs for between-study variances broadly agree across models, but estimates of between-study correlation coefficients remain imprecise with 95% CIs that nearly cover the permissible range defined in the prior distributions. Although the correlation between SBP and DBP is estimated to be highly positive with reasonable precision, the correlations for other pairs of outcomes vary across different assumed structures, and the estimated values have low precision. In the correlation plots in Appendix C, the multivariate analyses show a clearly linear positive correlation between the effects sizes for SBP and DBP; however, no pattern is discernible for correlations between other pairs of outcomes. This is likely due in part to more trials assessing both SBP and DBP than any other pair of outcomes (Table II). However, it is noteworthy that estimates for between-study variances are consistent across models.

8. Discussion

We have extended Bayesian multivariate meta-analysis methods to problems with more than two outcomes, allowing for the fact that not all outcomes of interest are necessarily reported by all studies. The proposed marginal likelihood approach makes use of available data from studies and avoids artificial imputation of missing treatment effect estimates and their associated within-study variance.

Our objective is to take into account the dependency between outcomes and to produce an estimate of the variance–covariance matrix, particularly in the challenging situation of a number of studies that is small relative to the number of outcomes. We are unable to provide concrete guidance on what we mean by 'small'. In the univariate random-effects meta-analysis, we might consider six studies to be a small number to estimate the two parameters in the model. Our model has $\left(p + \sum_{i=1}^{p} i\right)$ unknown parameters for p outcomes of interest. If three studies per parameter were to be considered a minimum reasonable sample size, we would require $3\left(p + \sum_{i=1}^{p} i\right)$ studies. This suggests that to analyse the unstructured model, we would seek 15 studies for a bivariate problem, 27 for a trivariate problem and 42 for a four-outcome problem such as the one we address in our example (comparing well with the 47 we had).

We considered a Bayesian approach because of the extra flexibility of incorporating prior information or assumptions. An alternative approach has been proposed that avoids the estimation of the variance–covariance matrix by addressing both within-study and between-study variance structures using a single matrix [11]. Using this approach, the number of parameters is reduced and the within-study correlations are not needed. However, the alternative model does not estimate the between-study variance τ^2. Although the estimate of overall variance is close to τ^2 when the between-study variance is larger compared with the within-study variance, this is not the case otherwise. A recently proposed approach using robust estimation methods appears to be suitable only for a large numbers of studies [5], which is an uncommon situation in medicine.

We separate the variance–covariance matrix so that variances and correlations are modelled separately, offering more flexibility than the use of a Wishart prior distribution. Notably, the separation strategy allows for more natural incorporation of prior information from expert belief or data-based distributions. Indeed, a particular disadvantage of the Wishart prior distribution is that it does not provide a straightforward option for incorporating informative prior information. Recent work has formulated informative prior distributions for heterogeneity parameters based on the meta-analyses in the *Cochrane Database of Systematic Reviews* [26, 27]. These priors were proposed for use in a univariate meta-analysis and now can be used in a multivariate setting by using the separation strategy. In our example, the two

Table III. Comparison of results of Bayesian multivariate meta-analyses (Cholesky decomposition) with various structured between-study variance–covariance matrix for acute stroke data.

	Structure I		Structure II		Structure III		Structures I and II		Structures I and III		Structures II and III	
Parameter of interest	Posterior median [95% CI]	Posterior SD	Posterior median [95% CI]	Posterior SD	Posterior median [95% CI]	Posterior SD	Posterior median [95% CI]	Posterior SD	Posterior median [95% CI]	Posterior SD	Posterior median [95% CI]	Posterior SD
μ_1	1.09 [0.94, 1.22]	0.06	1.06 [0.94, 1.23]	0.07	1.05 [0.90, 1.21]	0.07	1.05 [0.93, 1.20]	0.07	1.09 [0.93, 1.23]	0.09	1.05 [0.91, 1.21]	0.08
μ_2	1.09 [0.95, 1.20]	0.07	1.06 [0.91, 1.24]	0.08	1.04 [0.89, 1.19]	0.07	1.04 [0.92, 1.23]	0.07	1.07 [0.92, 1.25]	0.09	1.04 [0.89, 1.21]	0.08
μ_3	−3.91 [−5.7, 1.89]	0.96	−3.96 [−5.97, −1.90]	1.02	−3.97 [−6.00, −1.97]	1.01	−4.01 [−5.89, −2.06]	0.98	−3.77 [−6.12, −1.71]	1.08	−4.00 [−5.97, 2.04]	0.99
μ_4	−2.53 [−3.70, −1.18]	0.66	−2.50 [−3.88, −1.16]	0.68	−2.60 [−3.92, −1.19]	0.70	−2.52 [−3.87, −1.21]	0.68	−2.44 [−3.97, −0.93]	0.73	−2.57 [−3.95, −1.22]	0.69
τ_1^2	0.000 [0.000, 0.033]	0.01	0.006 [0.000, 0.022]	0.02	0.005 [0.000, 0.073]	0.02	0.005 [0.000, 0.053]	0.02	0.000 [0.000, 0.069]	0.02	0.006 [0.000, 0.091]	0.03
τ_2^2	0.000 [0.000, 0.033]	0.01	0.014 [0.000, 0.13]	0.03	0.008 [0.000, 0.115]	0.03	0.005 [0.000, 0.053]	0.02	0.000 [0.000, 0.069]	0.02	0.021 [0.000, 0.157]	0.04
τ_3^2	13.57 [4.70, 33.99]	7.94	13.26 [4.29, 33.65]	7.63	16.28 [6.11, 38.86]	8.93	13.19 [4.50,34.43]	7.84	15.28 [5.23, 35.14]	7.83	14.61 [5.509, 36.45]	8.02
τ_4^2	8.33 [3.45, 18.34]	4.00	7.85 [3.29, 16.87]	3.52	9.70 [4.544, 22.69]	4.36	7.91 [3.43, 16.95]	3.55	9.21 [4.01, 18.16]	3.88	8.722 [3.803, 18.39]	3.78
ρ_{12}	−0.07 [−0.99, 0.99]	0.69	−0.001 [−0.94, 0.94]	0.59	0.72 [0.039, 1.00]	0.31	0.014 [−0.95,0.94]	0.57	0.78 [0.043, 1.00]	0.30	0.523 [0.03, 0.98]	0.29
ρ_{13}	−0.04 [−1.00, 0.99]	0.77	0.057 [−0.74, 0.78]	0.40	0.72 [0.05, 0.99]	0.29	0.05 [−0.83,0.80]	0.44	0.76 [0.05, 1.00]	0.26	0.472 [0.031, 0.876]	0.24
ρ_{14}	−0.04 [−1.00, 0.99]	0.78	0.050 [−0.70, 0.75]	0.39	0.74 [0.08, 1.00]	0.27	0.04 [−0.81,0.77]	0.016	0.76 [0.08, 1.00]	0.26	0.421 [0.025, 0.856]	0.24
ρ_{23}	0.025 [−0.99, 0.96]	0.63	0.057 [−0.74, 0.78]	0.40	0.77 [0.17, 0.99]	0.23	0.05 [−0.83,0.80]	0.44	0.78 [0.15, 0.99]	0.22	0.472 [0.031, 0.875]	0.24
ρ_{24}	−0.08 [−0.95, 0.96]	0.62	0.05 [−0.70, 0.75]	0.39	0.69 [0.16, 0.99]	0.24	0.04 [−0.81,0.77]	0.43	0.71 [0.16, 1.00]	0.23	0.421 [0.025, 0.856]	0.24
ρ_{34}	0.86 [0.45, 0.99]	0.14	0.82 [0.37, 0.98]	0.16	0.90 [0.60, 0.99]	0.10	0.822 [0.36,0.97]	0.16	0.89 [0.55, 0.99]	0.11	0.852 [0.496, 0.984]	0.13
$\bar{D}$	464.31		459.97		462.55		461.44		464.22		459.90	
pD	34.81		40.27		37.72		37.62		34.81		40.15	
DIC	499.18		500.24		500.27		499.06		499.52		500.05	
Run time (s)	15 028		38 490		15 015		29 161		14 653		29 739	

Results are based on 100 000 MCMC samples with first 10 000 iterations discarded as burn-in and thinning to every 10th sample. DIC, deviance information criterion.

Statistics in Medicine

dichotomous outcomes consist of one objective outcome and one semi-objective outcome, with comparisons being between pharmacological interventions versus placebo. According to Turner *et al.* [27], appropriate prior distributions for the heterogeneity variance in these situations are

$$\tau_1^2 \sim \mathrm{LN}(-4.06, 1.52^2),$$
$$\tau_2^2 \sim \mathrm{LN}(-3.02, 1.85^2),$$

which might be used as data-based alternatives to the priors specified in Section 5.

We have considered two parameterizations of the variance–covariance matrix, Cholesky and spherical parameterization. The latter allows the incorporation of a prior assumption that the between-study correlation is positive, whereas the former additionally facilitates the assumption of homogeneous between-study correlations. Both parameterizations can be, in principle, extended to more than four outcomes.

We have considered the separation strategy for both structured and unstructured variance–covariance matrices; structured matrices can be particularly appealing when the amount of data is limited. In our example, reducing complexity in the variance–covariance matrix increased precision of parameter estimates for between-study variances. However, this does not necessarily indicate that an arbitrarily large number of outcomes can safely be included in a multivariate setting. If there are many more than four outcomes, for example, we anticipate that a number of problems can arise. Computation is expensive both in preparing the prior distributions and in running the MCMC algorithms; and estimation becomes more difficult when there are many more parameters in the models. The question of whether or not multivariate meta-analysis should routinely be pursued for multiple outcomes therefore remains a question for further research.

Which method would we recommend in practice? In our example, the Wishart prior was the quickest method, taking only 60 s to run. However, the estimation of the between-study variance–covariance matrix can be dependent on the prior, as shown in our sensitivity analysis (Figure 1), and it is difficult to specify deliberately informative prior distributions. We prefer the separation strategy using decomposition techniques, where we are able to specify priors with greater flexibility. Between Cholesky and spherical decompositions, we have found no clear advantage of one over another, other than the practical difficulty of implementing structure II with a spherical decomposition.

The proposed Bayesian multivariate methods allow simultaneous synthesis of any number of outcomes from multiple studies. In our example analyses, multivariate methods outperformed univariate methods, suggested by a slightly higher precision and a preferable balance of model fit against complexity (as measured by DIC). Recent work has demonstrated notable advantages of multivariate methods over univariate methods in the presence of selective nonreporting of a subset of the outcomes [49]. The methods we propose here merit more extensive examination, in particular to assess how many studies are needed for robust estimation in a multivariate setting as the number of outcomes increases. With the separation strategy, both noninformative and informative priors are now feasible in the context of multivariate meta-analysis.

Appendix A. Derivation of Cholesky factors

A.1. Derivation of Cholesky factors for an unstructured 4×4 *correlation matrix*

We write the unstructured correlation matrix as

$$R = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} \\ \cdot & 1 & \rho_{23} & \rho_{24} \\ \cdot & \cdot & 1 & \rho_{34} \\ \cdot & \cdot & \cdot & 1 \end{bmatrix}, \tag{A.1}$$

where $\cdot$ in the nondiagonal entries represent the symmetric correlation coefficients specified in the upper triangle. We use Cholesky decomposition to write $R = L^{\mathrm{T}}L$, where L is a $p \times p$ upper-triangular matrix:

$$L = \begin{bmatrix} L_{11} & 0 & 0 & 0 \\ L_{12} & L_{22} & 0 & 0 \\ L_{13} & L_{23} & L_{33} & 0 \\ L_{14} & L_{24} & L_{34} & L_{44} \end{bmatrix}. \tag{A.2}$$

Statistics in Medicine

The correlation matrix can be written in terms of L_{ij} $(i = 1, 2, 3, 4;\ i \leqslant j)$:

$$R = \begin{bmatrix} L_{11}^2 & L_{11}L_{12} & L_{11}L_{13} & L_{11}L_{14} \\ \cdot & L_{12}^2 + L_{22}^2 & L_{12}L_{13} + L_{22}L_{23} & L_{12}L_{14} + L_{22}L_{24} \\ \cdot & \cdot & L_{13}^2 + L_{23}^2 + L_{33}^2 & L_{13}L_{14} + L_{23}L_{24} + L_{33}L_{34} \\ \cdot & \cdot & \cdot & L_{14}^2 + L_{24}^2 + L_{34}^2 + L_{44}^2 \end{bmatrix}. \tag{A.3}$$

After such a Cholesky decomposition, a prior distribution needs to be placed on the elements L_{ij} instead of the correlation coefficients ρ_{ij} $(i = 1, \ldots, k;\ i < j)$. This requires knowledge of plausible intervals for the L_{ij} terms, which can be obtained by setting (A.1) equal to (A.3). First, we have $L_{11}^2 = 1$, which implies that $L_{11} = 1$ to obtain a one-to-one mapping. Then,

$$\begin{aligned} \rho_{12} &= L_{12}, \\ \rho_{13} &= L_{13}, \\ \rho_{14} &= L_{14}. \end{aligned}$$

Because these three L_{ij} are equal to correlation coefficients ρ_{1j}, their plausible intervals are

$$L_{12}, L_{13}, L_{14} \in [-1, 1]. \tag{A.4}$$

We then derive the plausible intervals for other elements using the conditions that $\mathbf{L}_k^{\mathrm{T}}\mathbf{L}_k = 1$, where $\mathbf{L}_k$ denotes the kth column in matrix (A.2).

From $L_{12}^2 + L_{22}^2 = 1$, we have

$$L_{22} = \sqrt{1 - L_{12}^2}. \tag{A.5}$$

Similarly,

$$L_{33} = \sqrt{1 - L_{13}^2 - L_{23}^2} \tag{A.6}$$

and

$$L_{44} = \sqrt{1 - L_{14}^2 - L_{24}^2 - L_{34}^2}, \tag{A.7}$$

where L_{23}^2, L_{24}^2 and L_{34}^2 are given as follows. Because $L_{13}^2 + L_{23}^2 + L_{33}^2 = 1$ and L_{23}^2 must be greater than or equal to 0, we have $L_{23}^2 = 1 - L_{13}^2 - L_{33}^2 \leqslant 1 - L_{13}^2$, so the plausible interval for L_{23} is

$$L_{23} \in \left[-\sqrt{1 - L_{13}^2}, \sqrt{1 - L_{13}^2}\right], \tag{A.8}$$

and in a similar way, we derive

$$L_{24} \in \left[-\sqrt{1 - L_{14}^2}, \sqrt{1 - L_{14}^2}\right]. \tag{A.9}$$

Finally, from $L_{14}^2 + L_{24}^2 + L_{34}^2 + L_{44}^2 = 1$,

$$L_{34}^2 = 1 - L_{14}^2 - L_{24}^2 - L_{44}^2 \leqslant 1 - L_{14}^2 - L_{24}^2,$$

so we have

$$L_{34} \in \left[-\sqrt{1 - L_{14}^2 - L_{24}^2}, \sqrt{1 - \rho_{14}^2 - L_{24}^2}\right]. \tag{A.10}$$

Uniform prior distributions are placed on each L_{ij} $(i = 1, \ldots, 4;\ i < j)$ within the associated plausible intervals given by (A.4) and (A.8)–(A.10), and L_{ii} $(i = 2, 3, 4)$ is given by (A.5), (A.6) and (A.7).

A.2. Derivation of Cholesky factors for a structured 4×4 *correlation matrix*

We assume a structure for the correlation matrix in which two pairs of correlation are identical as follows:

$$R_s = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \cdot & 1 & \rho_2 & \rho_3 \\ \cdot & \cdot & 1 & \rho_4 \\ \cdot & \cdot & \cdot & 1 \end{bmatrix}. \tag{A.11}$$

We set (A.11) equal to (A.3), and again set $L_{11}^2 = 1$, which implies that $L_{11} = 1$. We then write the correlation coefficients in terms of the L_{ij} $(i = 1, 2, 3, 4;\ i \leqslant j)$ as follows:

$$\begin{aligned} &\rho_1 = L_{12}, \\ &\rho_2 = L_{13} = L_{12}L_{13} + L_{22}L_{23} \Rightarrow \rho_2 = \frac{L_{22}L_{23}}{1 - L_{12}} = \frac{\sqrt{1 - L_{12}^2}}{1 - L_{12}} L_{23} = \frac{\sqrt{1 - \rho_1^2}}{1 - \rho_1} L_{23}, \\ &\rho_3 = \frac{\sqrt{1 - \rho_1^2}}{1 - \rho_1} L_{24}, \\ &\rho_4 = L_{13}L_{14} + L_{23}L_{24} + L_{33}L_{34} = \rho_1\rho_3 + + L_{23}L_{24} + L_{34}\sqrt{1 - \rho_2^2 - L_{23}^2}. \end{aligned} \tag{A.12}$$

It is obvious that

$$L_{12} \in [-1, 1]. \tag{A.13}$$

Because

$$\begin{aligned} &L_{13}^2 + L_{23}^2 + L_{33}^2 = 1 \\ &\Rightarrow \left(\frac{\sqrt{1 - L_{12}^2}}{1 - L_{12}} L_{23} \right)^2 + L_{23}^2 + L_{33}^2 = 1 \\ &\Rightarrow \left(\frac{\sqrt{1 - L_{12}^2}}{1 - L_{12}} L_{23} \right)^2 + L_{23}^2 \leqslant 1 \\ &\Rightarrow \frac{1 - L_{12}^2 + 1 - 2L_{12} + L_{12}^2}{(1 - L_{12})^2} L_{23}^2 \leqslant 1 \\ &\Rightarrow L_{23}^2 \leqslant \frac{1 - L_{12}}{2}, \end{aligned}$$

then

$$L_{23} \in \left[-\sqrt{\frac{1 - L_{12}}{2}}, \sqrt{\frac{1 - L_{12}}{2}} \right]. \tag{A.14}$$

Similarly, we have

$$L_{24} \in \left[-\sqrt{\frac{1 - L_{12}}{2}}, \sqrt{\frac{1 - L_{12}}{2}} \right]. \tag{A.15}$$

Because

$$
\begin{aligned}
&L_{14}^2 + L_{24}^2 + L_{34}^2 + L_{44}^2 = 1 \\
&\Rightarrow \left(\frac{\sqrt{1-L_{12}^2}}{1-L_{12}}\right)^2 L_{24}^2 + L_{24}^2 + L_{34}^2 \leqslant 1 \\
&\Rightarrow \frac{1-L_{12}^2+1-2L_{12}+L_{12}^2}{(1-L_{12})^2} L_{24}^2 + L_{34}^2 \leqslant 1 \\
&\Rightarrow L_{34}^2 \leqslant 1 - \frac{2}{(1-L_{12})} L_{24}^2,
\end{aligned}
$$

then

$$
L_{34} \in \left[-\sqrt{1-\frac{2}{(1-L_{12})}L_{24}^2}, \sqrt{1-\frac{2}{(1-L_{12})}L_{24}^2}\right]. \tag{A.16}
$$

Uniform priors are placed on L_{12}, L_{23}, L_{24} and L_{34} over the plausible intervals given in (A.13)–(A.16), respectively.

Appendix B. WinBUGS code

B.1. Code for univariate model using uniform prior for heterogeneity parameters

Data:

y[]: vector of all outcome data reported from all studies.
outcomes[]: vector of outcome categories for data in corresponding entry of y[] (values are 1, 2, 3 or 4 in our example).
m: total number of outcome data points (length of y[]).
p: number of outcomes.

```
model
  {
    for (i in 1:m) {
      y[i] ~ dnorm(theta[i, outcome[i]], insigma[i])            # likelihood
      insigma[i] <-1/sigma[i]                                   # within-study precision
      theta[i, outcome[i]] ~ dnorm(mu[outcome[i]], intau[outcome[i]])   # random effects
    }
    for (i in 1:p) {
      mu[i] ~ dnorm(0.00000E+00, 1.00000E-04)      # overall treatment effects
      tau[i] <-pow(sd[i], 2)                       #heterogeneity parameters
      intau[i] <-1/tau[i]
    }
    sd[1] ~ dunif(0.00000E+00, 2)                  # outcomes 1 and 2 are dichotomous
  outcomes
    sd[2] ~ dunif(0.00000E+00, 2)
    sd[3] ~ dunif(0.00000E+00, 10)                 # outcomes 3 and 4 are continuous
  outcomes
    sd[4] ~ dunif(0.00000E+00, 10)
    eform.mu[1] <-exp(mu[1])
    eform.mu[2] <-exp(mu[2])
  }
```

B.2. Code for multivariate model using Wishart prior for variance–covariance matrix

Data:

y[]: vector of all outcome data reported from all studies. The first NU data points in this vector record the outcome information from studies reporting only one outcome. The remaining data points (y[NU + 1 : N]) record the outcome data from studies reporting more than one outcome.
m: total number of outcome data points (length of y[])
p: number of outcomes
N: number of studies
NU: number of studies reporting only one outcome
study[]: vector (length m) recording study index for data in corresponding entry of y[]
pos[]: vector (length N) recording starting position in y[] of outcome data for each study
len[] vector (length N) recording number of outcomes reported by each study

NB: The outcome data for study i is recorded in y[pos[i]: pos[i]+len[i]]

Sigma[]: vector recording within-study precision matrix for all studies
pos2[] vector (length N) recording the starting position within Sigma[] of entries of variance–covariance matrix for each study
len2[] vector (length N) recording the numbers of entries within Sigma[] that correspond to the variance–covariance matrix terms for each study

NB: The within-study precision matrix for study i is recorded in Sigma[pos2[i],pos2[i]+len2[i]].

X[,]: design matrix of indicator variables (dimension $N \times p$) indicating which outcomes are reported by each study. Specifically, vector $X[i,]$ specifies which outcomes are reported by study i. For example, if $m = 4$ and study i reports only outcome 2, then $X[i,] = (0, 1, 0, 0)$.

```
model
  {
    for (i in 1:NU) {
      y[i] ~ dnorm(theta[i], sigma[i]) # models for studies that reported only one outcome
    }
    for (i in NU + 1:N) {
      y[pos[i]:pos[i] + len[i] − 1] ~ dmnorm(theta[pos[i]:pos[i] +
      len[i] − 1], sigma[pos2[i]:pos2[i] + len2[i] - 1]) # models for more than one outcomes
    }
    # transforms random effects using design matrix
     for (i in 1:m) {
      theta[i] < −inprod(X[i, ], pre.theta[study[i], ])
    }
    for (i in 1:N) {
      pre.theta[i, 1:p] ~ dmnorm(mu[1:p], intau[1:p, 1:p])
    }
    for (i in 1:p) {
      mu[i] ~ dnorm(0.00000E+00, 0.001)          # overall treatment effects
    }
    intau[1:p, 1:p] ~ dwish(Omega[1:p, 1:p], p)    # Wishart priors
    tau[1:p, 1:p] < −inverse(intau[1:p, 1:p])
    eform.mu[1] < −exp(mu[1])
    eform.mu[2] < −exp(mu[2])
    for (i in 1:p − 1) {
      for (j in i + 1:p) {
        r[i, j] <- tau[i, j]/sqrt(tau[i, i] * tau[j, j])
      }
    }
  }
```

Statistics in Medicine

B.3. Code for multivariate model using Cholesky decomposition prior for variance–covariance matrix

```
model
  {
    for (i in 1:NU) {
      y[i] ~ dnorm(mu[i], sigma[i])
    }
    for (i in NU + 1:N) {
      y[pos[i]:pos[i] + len[i] − 1] ~ dmnorm(mu[pos[i]:pos[i] +
        len[i] − 1], sigma[pos2[i]:pos2[i] + len2[i] − 1])
    }
    for (i in 1:m) {
      mu[i] <- inprod(X[i, ], theta[study[i], ])
    }
    for (i in 1:N) {
      theta[i, 1:p] ~ dmnorm(beta[1:p], intau[1:p, 1:p])
    }
    for (i in 1:p) {
      beta[i] ~ dnorm(0.00000E+00, 0.001)
      tau[i, i] < −pow(sd[i], 2)
    }
    sd[1] ~ dunif(0.00000E+00, 2)
    sd[2] ~ dunif(0.00000E+00, 2)
    sd[3] ~ dunif(0.00000E+00, 10)
    sd[4] ~ dunif(0.00000E+00, 10)
    pi < −3.1415
    for (i in 1:3) {
      for (j in (i + 1):4) {
        tau[i, j] < −rho[i, j] * sd[i] * sd[j]
        tau[j, i] < −tau[i, j]
      }
    }
    L12 ~ dunif(−1, 1)
    L13 ~ dunif(−1, 1)
    L14 ~ dunif(−1, 1)
    L22.u ~ dunif(−1, 1)
    L22.s < −sqrt(1 − pow(L12, 2))
    L22 < −L22.s * L22.u
    L23.u ~ dunif(−1, 1)
    L23.s < −sqrt(1 − pow(L13, 2))
    L23 < −L23.s * L23.u
    L24.u ~ dunif(−1, 1)
    L24.s < −sqrt(1 − pow(L14, 2))
    L24 < −L24.u * L24.s
    L34.u ~ dunif(−1, 1)
    L34.s < −sqrt(1 − pow(L14, 2) − pow(L24, 2))
    L34 < −L34.u * L34.s
    L33.u ~ dunif(−1, 1)
    L33.s < −sqrt(1 − pow(L13, 2) − pow(L23, 2))
    L33 < −L33.s * L33.u
    L44.u ~ dunif(−1, 1)
    L44.s < −sqrt(1 − pow(L14, 2) − pow(L24, 2) − pow(L34, 2))
    L44 < −L44.u * L44.s
    rho[1, 2] < −L12
    rho[1, 3] < −L13
    rho[1, 4] < −L14
    rho[2, 3] < −L12 * L13 + L22 * L23
```

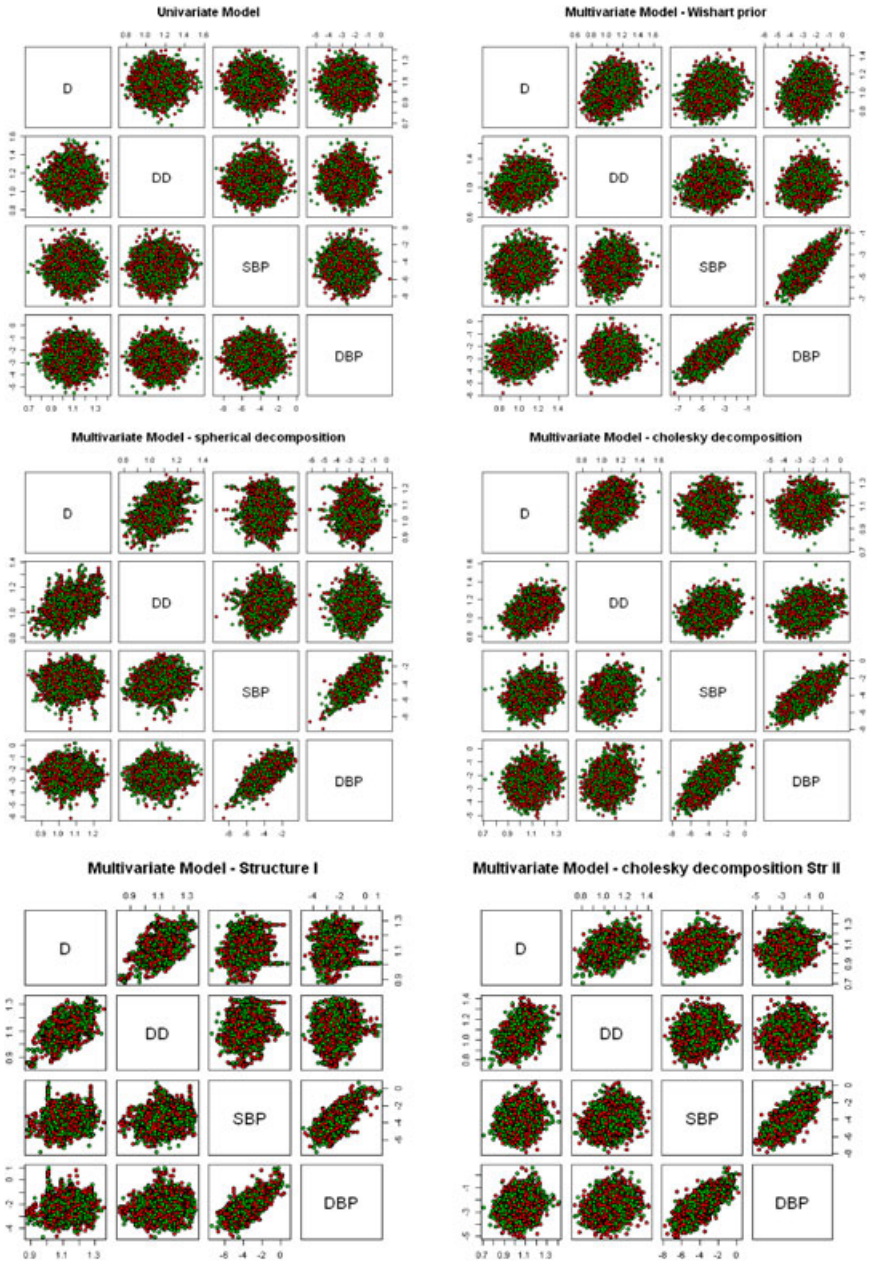
```
  rho[2, 4] < −L12 * L14 + L22 * L24
  rho[3, 4] < −L13 * L14 + L23 * L24 + L33 * L34
  for (i in 1:p − 1) {
    for (j in i + 1:p) {
      rho[j, i] < −rho[i, j]
    }
  }
  intau[1:p, 1:p] < −inverse(tau[1:p, 1:p])
  eform.beta[1] < −exp(beta[1])
  eform.beta[2] < −exp(beta[2])
  for (i in 1:p − 1) {
    for (j in i + 1:p) {
      r[i, j] < −tau[i, j]/sqrt(tau[i, i] * tau[j, j])
    }
  }
}
```

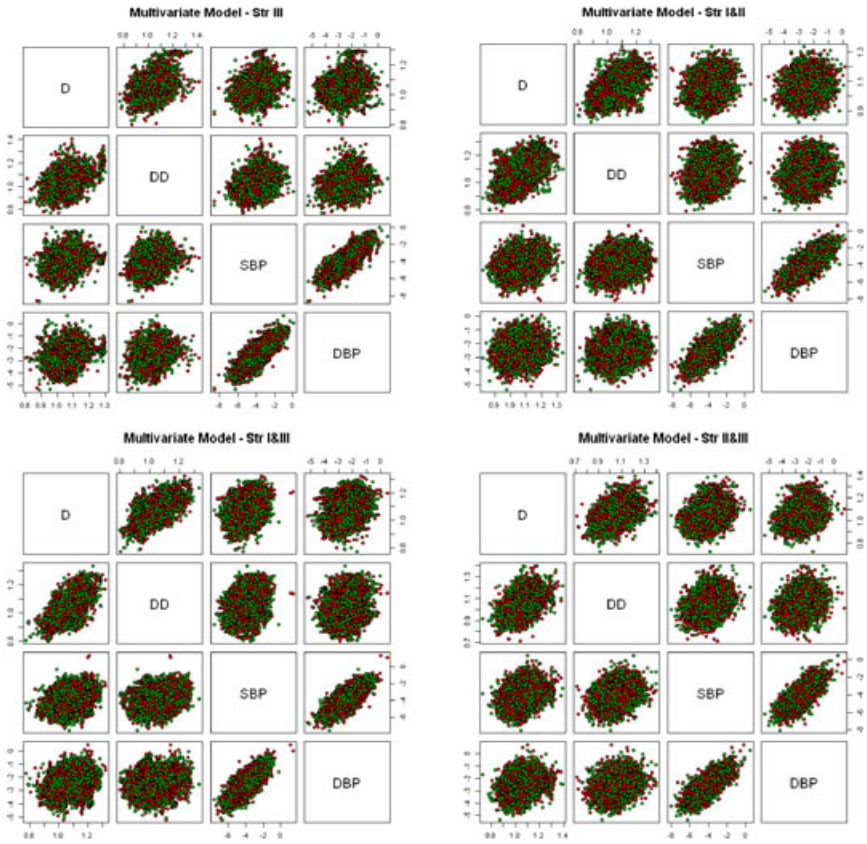
B.4. Code for multivariate model using spherical decomposition prior for variance–covariance matrix

```
model
  {
    for (i in 1:NU) {
      y[i] ∼ dnorm(theta[i], sigma[i])
    }
    for (i in NU + 1:N) {
      y[pos[i]:pos[i] + len[i] − 1] ∼ dmnorm(theta[pos[i]:pos[i] +
        len[i] −1], sigma[pos2[i]:pos2[i] + len2[i] − 1])
    }
    for (i in 1:m) {
      theta[i] < −inprod(X[i,], pre.theta[study[i], ])
    }
    for (i in 1:N) {
      pre.theta[i, 1:p] ∼ dmnorm(mu[1:p], intau[1:p, 1:p])
    }
    for (i in 1:p) {
      mu[i] ∼ dnorm(0.00000E+00, 0.001)
      tau[i, i] < −pow(sd[i], 2)
    }
    sd[1] ∼ dunif(0.00000E+00, 2)
    sd[2] ∼ dunif(0.00000E+00, 2)
    sd[3] ∼ dunif(0.00000E+00, 10)
    sd[4] ∼ dunif(0.00000E+00, 10)
    pi < −3.1415
    for (i in 1:3) {
      for (j in (i + 1):4) {
        tau[i, j] < −rho[i, j] * sd[i] * sd[j]
        tau[j, i] < −tau[i, j]
        g[j, i] < −0.00000E+00
        a[i, j] ∼ dunif(0.00000E+00, pi)
        rho[i, j] < −inprod(g[,i], g[,j])
      }
    }
    g[1, 1] < −1
    g[1, 2] < −cos(a[1, 2])
    g[2, 2] < −sin(a[1, 2])
    g[1, 3] < −cos(a[1, 3])
    g[2, 3] < −sin(a[1, 3]) * cos(a[2, 3])
    g[3, 3] < −sin(a[1, 3]) * sin(a[2, 3])
```

```
    g[1, 4] < −cos(a[1, 4])
    g[2, 4] < −sin(a[1, 4]) * cos(a[2, 4])
    g[3, 4] < −sin(a[1, 4]) * sin(a[2, 4]) * cos(a[3, 4])
    g[4, 4] < −sin(a[1, 4]) * sin(a[2, 4]) * sin(a[3, 4])
    eform.mu[1] < −exp(mu[1])
    eform.mu[2] < −exp(mu[2])
    intau[1:p, 1:p] < −inverse(tau[1:p, 1:p])
    for (i in 1:p − 1) {
      for (j in i + 1:p) {
        r[i, j] < −tau[i, j]/sqrt(tau[i, i] * tau[j, j])
      }
    }
}
```

Appendix C. Correlation plots for main effects posterior estimates





Acknowledgement

The authors were supported by the Medical Research Council [Unit Programme number U105285807].

References

1. Becker BJ. Multivariate meta-analysis. In *Handbook of Applied Multivariate Statistics and Mathematical Modeling*, Tinsley HEA, Brown SD (eds). Academic Press: San Diego, 2000; 499–525.
2. Berkey CS, Antczak-Bouckoms A, Hoaglin DC, Mosteller F, Pihlstrom BL. Multiple-outcomes meta-analysis of treatments for periodontal disease. *Journal of Dental Research* 1995; **74**(4):1030–1039. DOI: 10.1177/00220345950740040201.
3. Berkey CS, Anderson JJ, Hoaglin DC. Multiple-outcome meta-analysis of clinical trials. *Statistics in Medicine* 1996; **15**(5):537–557. DOI: 10.1002/(SICI)1097-0258(19960315)15:5<537::AID-SIM176>3.0.CO;2-S.
4. Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. *Statistics in Medicine* 1997; **16**(17):1965–1982. DOI: 10.1002/(SICI)1097-0258(19970915)16:17<1965::AID-SIM630>3.0.CO;2-M.
5. Hedges LV, Tipton E, Johnson MC. Robust variance estimation in meta-regression with depedent effect size estimates. *Research Synthesis Methods* 2010; **1**(1):39–65. DOI: 10.1002/jrsm.5.
6. Ishak K, Platt R, Joseph L, Hanley J. Impact of approximating or ignoring within-study covariances in multivariate meta-analyses. *Statistics in Medicine* 2008; **27**(5):670–686. DOI: 10.1002/sim.2913.
7. Jackson D, White IR, Thompson SG. Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. *Statistics in Medicine* 2010; **29**(12):1282–1297. DOI: 10.1002/sim.3602.
8. Nam IS, Mengersen K, Garthwaite P. Multivariate meta-analysis. *Statistics in Medicine* 2003; **22**(14):2309–2333. DOI: 10.1002/sim.1410.
9. 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**(1):78–97. DOI: 10.1002/sim.2524.
10. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR. Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Medical Research Methodology* 2007; **7**:3. DOI: 10.1186/1471-2288-7-3.
11. 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**(1):172–186. DOI: 10.1093/biostatistics/kxm023.
12. Sohn SY. Multivariate meta-analysis with potentially correlated marketing study results. *Naval Research Logistics* 2000; **47**(6):500–510. DOI: 10.1002/1520-6750(200009)47:6<500::AID-NAV3>3.0.CO;2-Z.

13. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine* 2002; **21**(4):589–624. DOI: 10.1002/sim.1040.
14. Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2009; **172**(4):789–811. DOI: 10.1111/j.1467-985X.2008.00593.x.
15. Jackson D, Riley R, White IR. Multivariate meta-analysis: Potential and promise. *Statistics in Medicine* 2011; **30**(20):2481–2498. DOI: 10.1002/sim.4172.
16. Mavridis D, Salanti G. A practical introduction to multivariate meta-analysis. *Statistical Methods in Medical Research* 2012. DOI: 10.1177/0962280211432219. (Epub ahead of print).
17. Ma Y, Mazumdar M. Multivariate meta-analysis: a robust approach based on the theory of U-statistic. *Statistics in Medicine* 2011; **30**(24):2911–2929. DOI: 10.1002/sim.4327.
18. Cheung MWL. metaSEM: Meta-analysis: A Structural Equation Modeling Approach, 2011. R package version 0.6.0, Available from: http://courses.nus.edu.sg/course/psycwlm/internet/metaSEM/ (accessed 11 October 2012).
19. Gasparrini A. mvmeta: multivariate meta-analysis and meta-regression, 2011. R package version 0.2.3, Available from: http://cran.r-project.org/web/packages/mvmeta/ (accessed 11 October 2012).
20. White IR. Multivariate random-effects meta-regression: update to mvmeta. *The Stata Journal* 2011; **11**(2):255–270.
21. White IR. Multivariate random-effects meta-analysis. *The Stata Journal* 2009; **9**(1):40–56.
22. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS- a Bayesian modeling framework: concepts, structure, and extensibility. *Statistics in Computing* 2000; **10**(4):325–337. DOI: 10.1002/sim.1410.
23. White IR, Barrett JK, Jackson D, Higgins JPT. Estimation of consistency and inconsistency models in multiple treatments meta-analysis using data augmentation and multivariate meta-regression. *Research Synthesis Methods* 2012; **3**(2):111–125. DOI: 10.1002/jrsm.1045.
24. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association* 2006; **101**:447–459. DOI: 10.1198/016214505000001302.
25. Pinheiro JC, Bates DM. Unconstrained parameterisations for variance–covariance matrices. *Statistics in Computing* 1996; **6**(3):289–296. DOI: 10.1007/BF00140873.
26. Pullenayegum EM. An informed reference prior for between-study heterogeneity in meta-analyses of binary outcomes. *Statistics in Medicine* 2011; **30**(26):3082–3094. DOI: 10.1002/sim.4326.
27. Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *International Journal of Epidemiology* 2012; **41**(3):818–827. DOI: 10.1093/ije/dys041.
28. Geeganage C, Bath PM. Vasoactive drugs for acute stroke. *Cochrane Database of Systematic Reviews* 2010; **7**. DOI: 10.1002/14651858.CD002839.pub2.
29. Gavish B, Ben-Dov IZ, Bursztyn M. Linear relationship between systolic and diastolic blood pressure monitored over 24 h: assessment and correlates. *Journal of Hypertension* 2008; **26**(2):199–209. DOI: 10.1097/HJH.0b013e3282f25b5a.
30. Cooper HM, Hedges LV. *The Handbook of Research Synthesis*. Russell Sage Foundation: New York, 1994.
31. Lu G, Welton NJ, Higgins JPT, White IR, Ades AE. Linear inference for mixed treatment comparison meta-analysis: a two-stage approach. *Research Synthesis Methods* 2011; **2**(1):43–60. DOI: 10.1002/jrsm.1059.
32. Schervish M. *Theory of Statistics*. Springer-Verlag: New York, 1995.
33. Spiegelhalter D, Thomas A, Best N, Lunn D. Winbugs User Manual 2003. Available from: http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf (accessed 11 October 2012).
34. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine* 2005; **24**(15):2401–2428. DOI: 10.1002/sim.2112.
35. Barnard J, McCulloch R, Meng XL. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica* 2000; **10**(4):1281–1331.
36. Daniels MJ, Pourahmadi M. Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika* 2002; **89**(3):553–556. DOI: 10.1093/biomet/89.3.553.
37. Daniels MJ, Kass RE. Nonconjugate bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association* 1999; **94**(448):1254–1263. DOI: 10.1080/01621459.1999.10473878.
38. Leonard T, Hsu JS. Bayesian inference for a covariance matrix. *Annals of Statistics* 1992; **20**(4):1669–1696. DOI: 10.1214/aos/1176348885.
39. Pourahmadi M. Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika* 2000; **87**(2):425–435. DOI: 10.1093/biomet/87.2.425.
40. Pourahmadi M. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterization. *Biometrika* 1999; **86**(3):677–690. DOI: 10.1093/biomet/86.3.677.
41. Yang R, Berger JO. Estimation of a covariance matrix using the reference prior. *Annals of Statistics* 1994; **22**(3):1195–211. DOI: 10.1214/aos/1176325625.
42. Lu G, Ades A. Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics* 2009; **10**(4):792–805. DOI: 10.1093/biostatistics/kxp032.
43. Wei Y, Higgins JP. Estimating within-study covariances in multivariate meta-analysis with multiple outcomes. *Statistics in Medicine* 2012. DOI: 10.1002/sim.5679. (Epub ahead of print).
44. Sturtz S, Ligges U, Gelman A. R2WinBUGS: A Package for Running WinBUGS from R. *Journal of Statistical Software* 2005; **12**(3):1–16.
45. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Statistical Science* 1992; **7**(4):457–511.
46. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 2002; **64**(4):583–639. DOI: 10.1111/1467-9868.00353.

47. Higgins J, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* 1996; **15**(24):2733–2749. DOI: 10.1002/(SICI)1097-0258(19961230)15:24<2733::AID-SIM562>3.0.CO;2-0.
48. Lumley T. Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine* 2002; **21**(16):2313–2324. DOI: 10.1002/sim.1201.
49. Kirkham JJ, Riley RD, Williamson PR. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. *Statistics in Medicine* 2012; **31**(20):2179–2195. DOI: 10.1002/sim.5356.