



# Modeling between-trial variance structure in mixed treatment comparisons

GUOBING LU\*, AE ADES

*Department of Community Based Medicine, University of Bristol,  
Cotham House, Cotham Hill, Bristol BS6 6JL, UK  
guobing.lu@bristol.ac.uk*

## SUMMARY

In mixed treatment comparison (MTC) meta-analysis, modeling the heterogeneity in between-trial variances across studies is a difficult problem because of the constraints on the variances inherited from the MTC structure. Starting from a consistent Bayesian hierarchical model for the mean treatment effects, we represent the variance configuration by a set of triangle inequalities on the standard deviations. We take the separation strategy (Barnard *and others*, 2000) to specify prior distributions for standard deviations and correlations separately. The covariance matrix of the latent treatment arm effects can be employed as a vehicle to load the triangular constraints, which in addition allows incorporation of prior beliefs about the correlations between treatment effects. The spherical parameterization based on Cholesky decomposition (Pinheiro and Bates, 1996) is used to generate a positive-definite matrix for the prior correlations in Markov chain Monte Carlo (MCMC). Elicited prior information on correlations between treatment arms is introduced in the form of its equivalent data likelihood. The procedure is implemented in a MCMC framework and illustrated with example data sets from medical research practice.

**Keywords:** Bayesian hierarchical model; Between-trial heterogeneity; Equivalent prior sample method; Evidence consistency; Markov chain Monte Carlo; Mixed treatment comparison meta-analysis; Triangle inequality.

## 1. INTRODUCTION

Meta-analysis of mixed treatment comparisons (MTC) (Lu and Ades, 2004), or network meta-analysis (Lumley, 2002), synthesizes comparative evidence on multiple treatments from randomized controlled trials (RCTs) available in a therapeutic area and has been increasingly used in medical decision making and health technology assessments. A brief review can be found in Sutton and Higgins (2008).

MTC models often assume, for conceptual and technical simplicity, that the between-trial variances are all equal (Higgins and Whitehead, 1996; Lumley, 2002), but in many situations, this is far from realistic. Frequently, particular pairs of treatments are pharmacologically similar to each other, but not to treatments in another class, and one might expect less between-trial variation where the treatments are more similar. For example, in trials of thrombolytic treatments following acute myocardial infarction, we might expect

\*To whom correspondence should be addressed.

a low between-trial variation, as all these treatments act physiologically in a similar way. Variations in patient group or protocol would, therefore, tend to affect all the treatments together, generating a high correlation in treatment effects and low between-trial variances. By contrast, in comparisons between these treatments and percutaneous transluminal coronary angioplasty, a surgical procedure, it would be reasonable to expect greater between-trial variation.

Spiegelhalter *and others* (2004) have argued that, where random effect models are used to summarize treatment effects, the predictive distribution of the effect in a new trial, for example,  $\delta_{XY}^{\text{new}} \sim N(d_{XY}, \sigma_{XY}^2)$ , may be a more appropriate representation of our belief about future efficacy than the mean effect  $d_{XY}$  (see also Ades *and others*, 2005). Since the variance of the predictive distribution depends critically on the population variance in effect size, it becomes more important that our estimate of the latter is unbiased. The assumption of homogeneous variance in these circumstances would often be difficult to defend.

However, modeling heterogeneous variances is a difficult task because the consistency of the MTC structure imposes some implicit constraints on the variances and covariances of random effects. Models in which the variances are assumed to be independent or treated as exchangeable random samples from a common distribution (Lu *and others*, 2007) are incompatible with the consistency relationships on which MTC models rely (see section 3.1).

In this paper, we propose an approach to modeling between-trial variance structures that is compatible with consistency assumptions and allows one to incorporate prior information on correlations between treatment arms, which may be an important consideration if data are sparse. Section 2 briefly describes exchangeability and consistency assumptions in Bayesian hierarchical modeling for MTC. Section 3 explains how the triangle inequalities are inherited from the requirement of evidence consistency and adopts a separation strategy to overcome the technical difficulty in embedding weakly informative prior distributions for the heterogeneity parameters. Section 4 discusses how to incorporate prior correlation information into a MTC heterogeneity variance model. We also apply our methods to 2 published data sets.

## 2. BAYESIAN HIERARCHICAL MODELS FOR MTC

### 2.1 MTC as incomplete data

Bayesian hierarchical models for MTC have been proposed by Lu and Ades (2004, 2006) and Lu *and others* (2007) under consistency conditions, extending ideas first proposed by Higgins and Whitehead (1996). Suppose that  $N$  RCTs have been conducted making mixed comparisons between  $K$  treatments. The comparisons are “mixed” in the sense that there may be only a subset of the treatments that are compared in each trial. Let

$$D_j = \{(r_{jk}, n_{jk}), k \in \mathcal{T}_j\}, \quad j = 1, \dots, N,$$

be the available data from the  $j$ th trial, where  $r_{jk}$  and  $n_{jk}$  are the numbers of events and cases, respectively, on treatment  $k$ , and  $\mathcal{T}_j$  is the set of treatments that are compared in trial  $j$ . We shall denote the baseline effect in trial  $j$  by  $\mu_j$  and denote for  $\forall h, k \in \mathcal{T}_j$  the relative effect of treatment  $k$  related to treatment  $h$  by  $\delta_{jhk}$ , on the log odds ratio scale. For narrative convenience, we also denote the set of treatments involved in all trials by  $\mathcal{T} = \cup \mathcal{T}_j$  and the set of trials that compare treatment pair  $(X, Y)$  by  $S_{XY}$ . Usually, the number of treatments in trial  $j$  is 2 or 3, leading to a very sparse data matrix. We thus need to develop a way of “borrowing strength” across both trials and MTC structure (Higgins and Whitehead, 1996). This requires a fully specified model that can match the MTC structure.

### 2.2 Borrowing strength from evidence consistency

Analogous to the well-known idea of missing at random in dealing with missing data, we may assume that the evidence about treatment contrasts from various sources is consistent in the sense that the “true”

relative effect of each treatment contrast remains the same over all trials under investigation whether or not it has been directly exposed to comparison in a particular trial. This leads to the basic assumption for a Bayesian hierarchical model as given below.

Assumption of exchangeability: For any treatment pair  $(X, Y)$ , the random relative effects  $\{\delta_{jXY}, j = 1, \dots, N\}$  are conditionally independent, given the true hyperparameters  $(d_{XY}, \sigma_{XY}^2)$ .

Note that although there are no direct comparisons between treatments  $(X, Y)$  in trials  $j \notin S_{XY}$ , these trials do contain indirect information about  $d_{XY}$ . For any treatment triple  $(X, Y, Z)$ , the definition of  $\delta$  on the log odds ratio (LOR) scale implies the following fundamental relationships in MTC.

$$\text{Transitivity relations: } \delta_{jXY} = \delta_{jZY} - \delta_{jZX}. \quad (2.1)$$

By taking expectations on both sides, Lu and Ades (2006) characterize evidence consistency with a system of conditions on the first-order moments called

$$\text{Consistency equations: } d_{XY} = d_{ZY} - d_{ZX} \quad \text{for } X \neq Y \neq Z \in \mathcal{T}. \quad (2.2)$$

These equations are essential for MTC modeling because they reflect the logical relationships between the population treatment contrasts and allow the model to borrow strength across the comparison networks. The unstructured hyperparameter set  $\mathcal{M} \equiv \{d_{bk}, k \neq b, b, k \in \mathcal{T}\}$  for population relative effects is now endowed with a coherent structure that effectively reduces the dimension of the parameter space.

### 2.3 Bayesian hierarchical models

Let  $b(j)$  be the specified baseline treatment in trial  $j$ . When there is no risk of confusion, we simply denote  $b(j)$  by  $b$ . Let  $p_{jk}$  be the probability of an event on treatment  $k$ . Under evidence consistency, a Bayesian hierarchical model in general settings can be written as follows:

$$\begin{aligned} r_{jk} &\sim \text{Bin}(p_{jk}, n_{jk}), k \in \mathcal{T}_j, \quad j = 1, \dots, N \\ p_{jk} &= \text{logit}^{-1}(\mu_j + X_{jk} \delta_{jbk}), \delta_{jbk} | d_{bk}, \sigma_{bk}^2 \sim N(d_{bk}, \sigma_{bk}^2) \\ d_{hk} &= d_{bk} - d_{bh} \quad \text{for } b \neq h \neq k \in \mathcal{T} \\ \text{Corr}(\delta_{jbh}, \delta_{jbk}) &= \gamma_{hk}^{(b)} \quad \text{for } b \neq h \neq k \in \mathcal{T}_j \\ \text{Priors for } \mu_j, d_{bk}, \sigma_{bk}^2, &\text{ and } \gamma_{hk}^{(b)}. \end{aligned} \quad (2.3)$$

The parameters are interpreted as follows. At the study level,  $\mu_j$  represents the specified baseline effect that is regarded as a nuisance parameter. As before,  $\delta_{jbk}$  is the treatment effect of treatment  $k$  related to the baseline  $b(j)$  on the logit scale. Here, using parameters  $(\mu_j, \delta_{jbk})$  rather than arm-effect parameters retain randomization in Bayesian meta-analysis (Smith *and others*, 1995; Thompson *and others*, 1997). The dummy variable  $X_{jk}$  is the indicator for baseline taking value 1 when  $k \neq b$  and value 0 when  $k = b$ . At the second level, while the hyperparameters  $d_{bk}$  and  $\sigma_{bk}^2$  are interpreted as before, the hyperparameters  $\gamma_{hk}^{(b)}$  describe the correlation between 2 treatment contrasts within a trial. These are only needed in a MTC model when there are multiarm trials on treatment triples  $(b, h, k)$ . For a full Bayesian analysis, all these hyperparameters require suitable prior distributions. While priors for  $d_{bk}$ , informative or noninformative, can be specified in a standard way (Smith *and others*, 1995; Thompson *and others*, 1997), the specification of variance priors is a difficult task which we address below.

### 3. PRIORS FOR HETEROGENEOUS VARIANCES

#### 3.1 Variance configuration in MTC

Denote the set of distinct hyperparameters for variances by  $\mathcal{V} \equiv \{\sigma_{bk}^2, b < k \in \mathcal{T}\}$ . By taking variances on both sides of the transitivity relations (2.1), for any treatment triple  $(b, h, k) \in \mathcal{T}$ , we have

$$\sigma_{hk}^2 = \sigma_{bh}^2 + \sigma_{bk}^2 - 2\gamma_{hk}^{(b)} \sigma_{bh} \sigma_{bk} \leq (\sigma_{bh} + \sigma_{bk})^2,$$

which in turn implies the second-order consistency conditions named as

$$\text{Triangle inequalities: } |\sigma_{bh} - \sigma_{bk}| \leq \sigma_{hk} \leq \sigma_{bh} + \sigma_{bk}. \quad (3.1)$$

Thus, the values of  $\sigma$ s form a configuration with any 3 standard deviations being able to configure a triangle (Figure 1). The mass of the joint distribution for  $\sigma$ s is concentrated on a cone contained in  $\mathfrak{R}_+^{K(K-1)/2}$ . When all  $\sigma$ s are equal, we obtain a homogeneity model with all  $\gamma$ s being 1/2 (Higgins and Whitehead, 1996). In this case, the triangle inequalities hold trivially. A MTC model satisfying both the first- and the second-order moment conditions would be more strictly a consistency model than one built on the first-order moment condition only. Also, under assumption of normality for contrast effects, the first- and second-order conditions would be sufficient to represent evidence consistency. However, the second-order moment constraint can cause serious difficulties in the specification of prior distributions for heterogeneous variances.

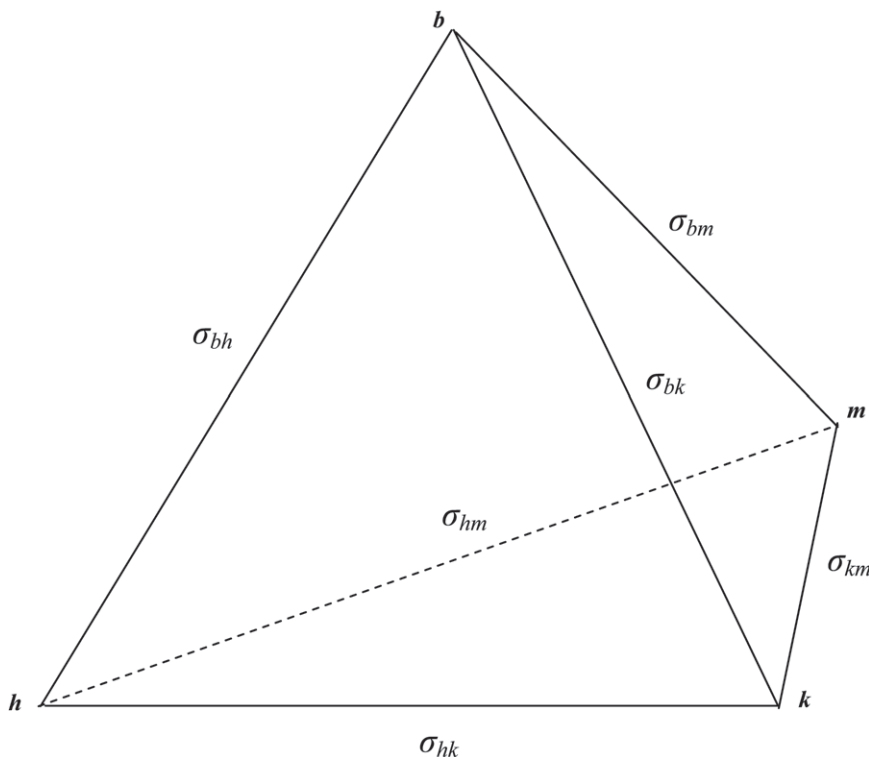


Fig. 1. Variance configuration for 4 treatment MTC: any 3 standard deviations configure a triangle with sds as sides.

Note that, once the values of  $\sigma$ s are given, the correlation coefficients  $\gamma_{hk}^{(b)}$ , when needed, can be determined by the cosine theory for triangles

$$\gamma_{hk}^{(b)} = \frac{\sigma_{bh}^2 + \sigma_{bk}^2 - \sigma_{hk}^2}{2\sigma_{bh}\sigma_{bk}} = \cos\langle \vec{bh}, \vec{bk} \rangle, \quad (3.2)$$

that is, each correlation coefficient is geometrically the cosine of the angle formed by sides  $bh$  and  $bk$ , which can be represented by lengths of the 3 sides in a triangle.

### 3.2 Ancillary variable method

We introduce some ancillary parameters as an instrument to guarantee triangle inequalities for  $\sigma$ s. As the random effect  $\delta_{jbk}$  is the LOR related to treatment pair  $(b, k)$  on the logit scale,  $\sigma_{bk}^2$  can be represented as

$$\sigma_{bk}^2 = \tau_b^2 + \tau_k^2 - 2\rho_{bk}\tau_b\tau_k, \quad (3.3)$$

where  $\tau_b^2$  and  $\tau_k^2$  are, respectively, variances of 2 random quantities  $\theta_{jb}$  and  $\theta_{jk}$  that can be interpreted as the random effects of corresponding treatment arms up to a common unknown constant, and  $\rho_{bk}$  is their correlation coefficient. Under this representation, it is easy to check that triangle constraints will be retained for the  $\sigma$ s. Let  $\Omega$  be the covariance matrix formed by  $\tau$ s and  $\rho$ s. Then, the prior distributions for the  $\sigma$ s under triangle constraints can be specified via a suitable distribution for  $\Omega$  and (3.3).

Special choices of  $\Omega$  can lead to particularly simplified models. For example, when all  $\tau$ s and all  $\rho$ s are equal, we obtain a homogeneity model with all  $\sigma$ s equal and all  $\gamma$ s in (3.2) equal 1/2. A slight extension allows  $\tau_k$  to vary freely but keeps  $\rho$ s being a constant, which may reflect our prior belief in the equality of interactions between treatment effects. Thus, when there is some prior knowledge or belief about the  $\Omega$  (see also example 2 in section 4), the ancillary parameter method may provide a useful way to incorporate such prior information into the model.

Note that, in general cases, the mapping from the ancillary parameter set  $\{\tau_k^2, \rho_{bk}, b < k\}$  to the variance parameter set  $\{\sigma_{bk}^2, b < k\}$  is not a one-to-one transformation. The role of ancillary parameters is to guarantee triangle constraints for the variances  $\sigma$ s.

### 3.3 Weakly informative prior distribution for $\Omega$

The basic requirement for specifying a prior distribution for  $\Omega$  is that it should be nonnegative definite. This problem, frequently encountered in Bayesian multivariate analyses for econometrics, psychology, education studies, and so forth, has been investigated by many statisticians, including Dempster (1972), Anderson *and others* (1987), Eaton and Olkin (1987), Leonard and Hsu (1992), Liu (1993), Yang and Berger (1994), Pinheiro and Bates (1996), Daniels and Kass (1999), Barnard *and others* (2000), and Sun and Sun (2005).

To achieve the flexibility we require in the prior for  $\Omega$ , we take the separation strategy (Barnard *and others*, 2000) to write

$$\Omega = \mathbf{V}^{1/2} \mathbf{R} \mathbf{V}^{1/2} \quad (3.4)$$

and specify independent priors for  $\mathbf{V}^{1/2}$  and  $\mathbf{R}$ , where  $\mathbf{V}^{1/2}$  is diagonal with elements  $\tau_k$ , and  $\mathbf{R}$  is the correlation matrix with nondiagonal elements  $\rho_{bk}$ . Possible choices of weakly informative prior for  $\tau_k$  include inverse gamma (IG) (Spiegelhalter *and others*, 2004), uniform (Warn *and others*, 2002), or half-t (Gelman, 2006).

To specify prior distributions for the correlation matrix  $\mathbf{R}$ , we use the Cholesky decomposition to write  $\mathbf{R} = \mathbf{L}^T \mathbf{L}$ , where  $\mathbf{L}$  is an  $K \times K$  upper-triangular matrix, and then obtain the weakly informative distribution for  $\mathbf{L}$  through the spherical parameterization (Pinheiro and Bates, 1996) as follows.

Let  $\mathbf{L}_k$  denote the  $k$ th column of the matrix  $\mathbf{L}$ , and  $\ell_k = (\ell_{k1}, \dots, \ell_{kk})^T$  the spherical coordinates of the first  $k$  elements of  $\mathbf{L}_k$ , given, for  $k = 2, \dots, K$ , by

$$\begin{aligned}\ell_{k1} &= \cos(\phi_{k2}) \\ \ell_{k2} &= \sin(\phi_{k2}) \cos(\phi_{k3}) \\ &\vdots \\ \ell_{k,k-1} &= \sin(\phi_{k2}) \sin(\phi_{k3}) \cdots \cos(\phi_{kk}) \\ \ell_{k,k} &= \sin(\phi_{k2}) \sin(\phi_{k3}) \cdots \sin(\phi_{kk}),\end{aligned}\tag{3.5}$$

with  $\ell_{11} = 1$ . To ensure uniqueness of the spherical parameterization, we must have  $\phi_{km} \in (0, \pi)$ , for  $m = 2, \dots, k$  (Pinheiro and Bates, 1996). Then, the  $(h, k)$  entry of  $\mathbf{R}$  can be represented as the inner product

$$\rho_{hk} = \mathbf{L}_h^T \mathbf{L}_k.\tag{3.6}$$

It is easy to check that the diagonal entries satisfy the condition  $\rho_{kk} = \mathbf{L}_k^T \mathbf{L}_k = \ell_k^T \ell_k = 1$ . For illustration, the  $3 \times 3$  upper-triangular matrix  $\mathbf{L}$  for Cholesky decomposition is given by

$$\mathbf{L} = \begin{pmatrix} 1 & \cos(\varphi_{21}) & \cos(\varphi_{31}) \\ 0 & \sin(\varphi_{21}) & \sin(\varphi_{31}) \cos(\varphi_{32}) \\ 0 & 0 & \sin(\varphi_{31}) \sin(\varphi_{32}) \end{pmatrix}\tag{3.7}$$

with the 3 coordinate parameters  $(\varphi_{21}, \varphi_{31}, \varphi_{32})$ , and the corresponding correlation matrix  $\mathbf{R}$  is

$$\mathbf{R} = \begin{pmatrix} 1 & \cos(\varphi_{21}) & \cos(\varphi_{31}) \\ * & 1 & \cos(\varphi_{21}) \cos(\varphi_{31}) + \sin(\varphi_{21}) \sin(\varphi_{31}) \cos(\varphi_{32}) \\ * & * & 1 \end{pmatrix}.\tag{3.8}$$

In the general case, all the angles in (3.5) may be given a uniform distribution over interval  $(0, \pi)$ , resulting in a general correlation matrix  $\mathbf{R}$  with  $\rho_{hk} \in (-1, 1)$ .

### 3.4 Positive correlation model

Within trials, each treatment is tested on the same population. It is therefore very likely that the random effects of treatment arms tend to be positively correlated so that  $\rho_{hk} \in (0, 1)$  for all  $h \neq k$ . In this case, all the angles in (3.5) should be given a uniform prior distribution on  $(0, \pi/2)$  instead of  $(0, \pi)$ .

As  $\phi \sim U(0, \pi/2)$ ,  $y = \cos(\phi)$  has the density  $p(y) = 2\pi^{-1}(1 - y^2)^{-1/2}$ ,  $y \in (0, 1)$ , which gives more mass to the value of  $y$  close to 1. However, we may specify prior distributions for  $y = \cos(\phi)$  directly by using the relation  $\sin(\phi) = (1 - y^2)^{1/2}$  for acute angle  $\phi$ . Thus, the spherical coordinates in (3.5) can be rewritten in terms of  $y$ s, and each  $y_i$  can be specified as beta  $(\alpha, \beta)$  distributions, where  $\alpha, \beta$  are some known constants or hyperparameters depending on our prior knowledge about the correlations between arm effects.

### 3.5 Example 1: Smoking cessation data

This data set (section 1 of the supplementary material available at *Biostatistics* online [<http://biostatistics.oxfordjournals.org/>]) comprises 24 trials on smoking cessation (Fiore and others, 1996), 22 two-arm and 2

three-arm trials, with 18 822 participants, making comparisons between 4 treatments: (A) no contact, (B) self-help, (C) individual counseling, and (D) group counseling. Meta-analysis for this data set was studied by Hasselblad (1998), Caldwell *and others* (2005), and Lu and Ades (2006). Although the heterogeneity is not too serious in this data set, we use it to illustrate our method. Considering that there are only 2 medium size 3-arm trials, in the following analysis, we ignore the multiarm effects and treat each 3-arm trial as 2 independent 2-arm trials.

We shall fit and compare 4 variance models, each constituted by combining model (1) with a particular between-trial variance structure:

- (i) HOM model: The usual homogeneity case in which all the between-trial variances are equal to  $\sigma^2$ , which is given a weakly informative prior, for example, IG distribution  $\text{IG}(10^{-3}, 10^{-3})$ .
- (ii) ID model: The unstructured heterogeneity model, assuming that  $\{\sigma_{XY}^2, X \prec Y\}$ , where “ $\prec$ ” means alphabetic priority, is independent and is given a weakly informative prior  $\text{IG}(10^{-3}, 10^{-3})$ . Because the triangle inequalities are not guaranteed, it is only a first-order consistency model.
- (iii) GC model: The general heterogeneity case under the consistency condition on the second-order moments where there is no prior information available about the correlations between treatment effects. In the Cholesky parameterization (3.5), we take  $\phi_{km} \sim U(0, \pi)$ .
- (iv) PC model: The specific case of the GC model with the further assumption that correlations between effects of treatment arms are all positive. In the Cholesky parameterization (3.5), we take  $\phi_{km} \sim U(0, \pi/2)$  and name the model as positive correlation (PC)-I. When taking  $y_{km} \equiv \cos(\phi_{km}) \sim \text{beta}(\alpha, \beta)$ , we name the model as PC-II when  $\alpha = \beta$  (note:  $\alpha = \beta = 1$  corresponds to  $y_{km} \sim U(0, 1)$ ), and as PC-III when  $\alpha \neq \beta$ .

We employ the Markov chain Monte Carlo (MCMC) method with the software WinBUGS (Spiegelhalter *and others*, 2001) to calculate posterior estimates of the parameters of interest. The code is given in section 2 of the supplementary material available at *Biostatistics* online (<http://biostatistics.oxfordjournals.org/>). Table 1 shows the posterior means for the basic treatment contrast parameters  $d_{AB}$ ,  $d_{AC}$ ,  $d_{AD}$ , and the standard deviation parameters  $\sigma$  and  $\sigma_{XY}$  under the above models. Figure 2 plots the posterior medians with 95% Bayesian credible intervals (CIs) for  $\sigma_{XY}$ 's under these models. The GC model and the 3 versions of PC model lead to very similar results, reflecting the insensitivity of the parameter estimates to the choice of weakly informative prior distribution for the Cholesky parameters for the correlation matrix  $\mathbf{R}$ . However, the ID model results in unstable estimates where data are sparse, without constraints of the triangle inequalities. The posterior mean of  $\sigma_{AD}$  under the ID model is as unrealistically high at 5.80 with a large standard deviation 12.86 and a very wide 95% CI (0.14, 27.92) compared with the posterior mean 0.95 with standard deviation 0.55 and 95% CI (0.15, 2.11) under the PC-I model, which highlights the role of triangle inequalities in “borrowing strength” for informing between-trial variances across treatment contrasts. Under the triangle inequalities, the values of  $\sigma_{AD}$  is bounded above by  $\min\{\sigma_{AX} + \sigma_{XD}, X = B, C\}$ . Table 1 also lists the means and standard deviations of the predictive distributions for relative treatment effects  $\delta_{XY}^{\text{new}}$  for  $X, Y \in \mathcal{T}$ , where for the same reason as above the ID model leads to a very flat predictive distribution for  $\delta_{AD}^{\text{new}}$ . The PC-I model seems particularly successful in limiting posterior uncertainty in prediction.

A useful reference measure for assessing adequacy and guiding selection of Bayesian hierarchical models is the deviance information criterion (DIC) proposed by Spiegelhalter *and others* (2002):  $\text{DIC} = \bar{D} + pD$ , where  $\bar{D}$  is the posterior mean of the Bayesian deviance measuring model fitness and  $pD$  is the “effective dimension” of the model serving as a penalty for the increased model complexity. The smallest  $\bar{D}$  value occurs in the ID model simply because it is the least restricted among the 4 models and has the lowest DIC. However, in a strict sense of evidence consistency, it does not qualify as a candidate model because it does not align with the theoretical assumption about evidence consistency. The PC-I and PC-II



Table 1. (Hassulblad data) Comparison of posterior mean under 4 variance models. The multi-arm trial effects on 2 three-arm trials have been ignored. Treatment effects are reported on the log odds scale. Results are based on 30 000 MCMC samples after 20 000 iterations burn-in. Some unrealistically high values under the ID model are underlined

Parameter of interest	Variance model (posterior mean/sd)					
	HOM	ID	GC: $\varphi \sim U(0, \pi)$	PC-I: $\varphi \sim U(0, \pi/2)$	PC-II: $\cos(\varphi) \sim U(0, 1)$	PC-III: $\cos(\varphi) \sim \text{beta}(5, 3)$
$d_{AB}$	0.523 (0.38)	0.38 (0.26)	0.540 (0.35)	0.53 (0.34)	0.54 (0.35)	0.53 (0.34)
$d_{AC}$	0.814 (0.23)	0.70 (0.23)	0.790 (0.24)	0.79 (0.24)	0.79 (0.24)	0.80 (0.24)
$d_{AD}$	1.167 (0.45)	1.05 (0.36)	1.182 (0.46)	1.19 (0.46)	1.18 (0.43)	1.18 (0.44)
$\sigma$	0.821 (0.18)					
$\sigma_{AB}$		0.36 (0.51)	0.71 (0.36)	0.71 (0.36)	0.74 (0.34)	0.71 (0.34)
$\sigma_{AC}$		0.90 (0.24)	0.88 (0.22)	0.88 (0.21)	0.88 (0.22)	0.88 (0.21)
$\sigma_{AD}$		<u>5.80</u> (12.86)	1.00 (0.54)	0.95 (0.55)	0.93 (0.45)	0.92 (0.48)
$\sigma_{BC}$		0.41 (1.04)	0.44 (0.34)	0.45 (0.35)	0.45 (0.33)	0.47 (0.34)
$\sigma_{BD}$		0.45 (1.42)	0.57 (0.57)	0.54 (0.64)	0.53 (0.53)	0.54 (0.56)
$\sigma_{CD}$		0.52 (0.52)	0.61 (0.59)	0.55 (0.62)	0.55 (0.51)	0.53 (0.51)
$\delta_{AB}^{\text{new}}$	0.518 (0.88)	0.37 (0.62)	0.54 (0.86)	0.53 (0.87)	0.54 (0.89)	0.53 (0.88)
$\delta_{AC}^{\text{new}}$	0.805 (0.83)	0.70 (0.97)	0.79 (0.94)	0.81 (0.95)	0.80 (0.95)	0.80 (0.94)
$\delta_{AD}^{\text{new}}$	1.163 (0.91)	1.02 ( <u>19.14</u> )	1.19 (1.24)	1.18 (1.11)	1.21 (1.47)	1.17 (1.12)
$\delta_{BC}^{\text{new}}$	0.295 (0.88)	0.35 (1.08)	0.25 (0.66)	0.27 (0.66)	0.26 (0.66)	0.27 (0.67)
$\delta_{BD}^{\text{new}}$	0.658 (0.92)	0.69 (1.12)	0.64 (0.96)	0.64 (0.82)	0.66 (1.25)	0.64 (0.89)
$\delta_{CD}^{\text{new}}$	0.368 (0.91)	0.35 (1.21)	0.40 (0.97)	0.39 (0.85)	0.41 (1.21)	0.38 (0.86)
$\bar{D}$	54.65	52.88	54.09	54.10	54.07	54.23
$pD$	43.64	42.44	43.26	43.10	43.13	43.12
DIC	98.29	95.32	97.35	97.20	97.20	97.35

models have similar second smallest DIC values, but the former gives better predictive distributions with smaller standard deviations. Thus, we would suggest the use of the PC-I model. We also performed a simulation study to examine performance of the PC-I model by running WinBUGS in R with the package R2WinBUGS (Gelman, 2007). See section 3 of the supplementary material available at *Biostatistics* online (<http://biostatistics.oxfordjournals.org>).

Note that, in the smoking cessation data, the heterogeneity between treatment contrasts is not very serious; hence, the HOM model might be acceptable because of its simplicity and the similarity in results to the PC models. However, although the data does not counter-indicate the homogeneity model, it is not plausible *a priori*, as one might expect the response to individual counseling (C) and group counseling (D) more “similar” to each other than to self-help (B).

#### 4. INCORPORATING PRIOR CORRELATION INTO A PC MODEL

##### 4.1 Elicitation of prior information about correlation

Researchers may have some prior information about between-treatment interaction or hold *a priori* certain beliefs about relations between treatment effects. Any sort of similarity in treatment protocols, biochemical or procedural, may generate higher or lower correlations between treatment effects. However, incorporating prior correlation information is difficult because the correlation coefficients between different treatment pairs are implicitly correlated according to a very complex joint distribution (Barnard *and others*, 2000).



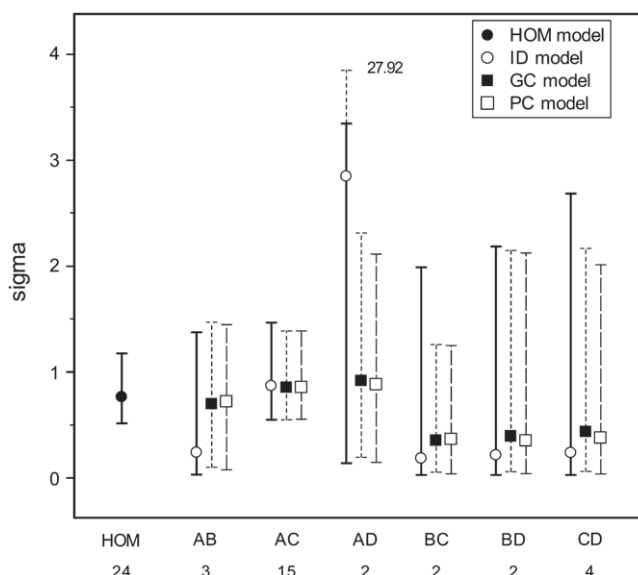


Fig. 2. Posterior medians with 95% Bayesian CIs for standard deviations under HOM-, ID-, GC- and PC models for smoking cessation data. The upper 2.5% quantile for  $\sigma_{AD}$  under ID model is as unrealistically high as 27.92, about 12 times of that under other models. The figures below case indices are the numbers of trials that make corresponding comparisons. Note that there are two 3-arm trials.

The ancillary variable method in section 3.2 with the Cholesky parameterization (3.5) can provide a way of introducing prior correlation information into a MTC model. In principle, elicitation variables are physically meaningful quantities and can be measured by procedures in which experts quantify their uncertainty on the basis of their expertise (Bedford and Cooke, 2001; O'Hagan *and others*, 2006). In MTC models, the parameters on which experts might be expected to have beliefs are  $\rho_{bk}$ , the correlation coefficients between treatment arms, rather than the angle parameters  $\phi_{kj}$ . Now suppose that, under the fundamental belief that all  $\rho$ s should be positive, the prior information about  $\rho$ s, including expert knowledge, can be quantified in the form of their marginal distributions, say means and standard deviations or medians and 95% CIs. An appealing choice of prior for each  $\rho_{bk}$  is the beta distribution with unknown parameters  $\alpha_{bk}, \beta_{bk}$  because “it can portray a good variety of opinions reasonably accurately” (O'Hagan *and others*, 2006, p. 125). We then apply the equivalent prior sample (EPS) information procedure (Winkler, 1967) in an inverse form to elicit experts' opinions and use the Cholesky parameterization to generate the joint prior distribution for  $(\rho_{bk}, b < k)$ , which has the sample space  $S^+$  of positive-definite matrices.

This procedure can be briefly described as follows:

- (i) Based on the experts' knowledge, choose the suitable data set:

$$D_{\text{EPS}} = \{(\alpha_{bk}, \beta_{bk}) : b < k\}, \quad (4.1)$$

where  $\alpha_{bk}$  are hypothetical samples from Binomial-type distribution with “sample size”  $\alpha_{bk} + \beta_{bk}$  and success probabilities  $\rho_{bk}$ , with the density proportional to

$$\rho_{bk}^{\alpha_{bk}} (1 - \rho_{bk})^{\beta_{bk}}.$$

Note that  $(\alpha_{bk}, \beta_{bk})$  are allowed to be any positive real numbers.

(ii) Use the functional relation given by (3.5) and (3.6), that is,

$$\boldsymbol{\rho} = f(\boldsymbol{\phi}), \quad \text{with } \rho_{bk} = f_{bk}(\boldsymbol{\phi}) = \mathbf{L}_b^T(\boldsymbol{\phi})\mathbf{L}_k(\boldsymbol{\phi}), \quad (4.2)$$

where  $\boldsymbol{\rho} = (\rho_{bk}, b < k)$  denotes the  $K(K-1)/2$  dimensional vector of correlation coefficients and  $\boldsymbol{\phi} = (\phi_{kj}, k = 2, \dots, K, 2 \leq j \leq k)$  the corresponding vector of angles in the Cholesky decomposition.

(iii) Specify a noninformative prior  $\pi(\boldsymbol{\phi})$ .

The above procedure allows us to incorporate prior information  $D_{\text{EPS}}$  on  $\rho_{bk}$ , even through  $\rho_{bk}$  is a functional parameter. This leads to the “posterior” joint distribution for the random vector  $\boldsymbol{\rho}$  on  $S_+$ , given the hypothetical data  $D_{\text{EPS}}$ , with density

$$\pi(\boldsymbol{\rho} | D_{\text{EPS}}) = g(f^{-1}(\boldsymbol{\rho})) \left| \frac{\partial}{\partial \boldsymbol{\rho}} f^{-1}(\boldsymbol{\rho}) \right|, \quad (4.3)$$

where  $g(\boldsymbol{\phi}) \equiv \pi(\boldsymbol{\phi} | D_{\text{EPS}}) \propto \prod_{b < k} f_{bk}(\boldsymbol{\phi})^{\alpha_{bk}} (1 - f_{bk}(\boldsymbol{\phi}))^{\beta_{bk}} \pi(\boldsymbol{\phi})$ .

An example for illustration of EPS method is given in section 4 of the supplementary material available at *Biostatistics* online (<http://biostatistics.oxfordjournals.org>). O’Hagan *and others* (2006) noted that methods for eliciting parameters of multivariate distributions are complex and not wholly satisfying. As our scheme guarantees a positive-definite correlation matrix, it is possible to use simpler methods for eliciting priors for scalar quantities. Respondents could be asked to give a “best” estimate of the correlation (between 0 and 1) and credible limits. O’Hagan *and others* (2006, chapter 6) describe a number of methods. The results can then be readily approximated as a beta distribution.

## 4.2 Example 2: Gastroesophageal reflux disease data

This data set consists of 41 RCTs on gastroesophageal reflux disease (GERD) with a total of 4637 participants, making mixed comparisons between 6 common therapies: (1) placebo, (2) prokinetic agents, (3)  $H_2$  receptor antagonists ( $H_2$ RA), (4)  $H_2$ RA double dose, (5) proton pump inhibitors (PPI), and (6) PPI double dose. Outcomes of these trials were observed at one or more follow-up times, varying with trials, of 4, 6, 8, or 12 weeks after commencement of treatment. For a detailed description of the data and graphical representation of the comparison networks refer to Goeree *and others* (1999) and Lu *and others* (2007).

To borrow strength both across MTC networks and across time points, Lu *and others* (2007) applied a set of piecewise exponential models on the log hazards with random effects of both treatments and time periods and briefly allowed an exchangeable half-t model to deal with the between-trial heterogeneity. Their model 8 was

$$\begin{aligned} p_{jku'u} &= 1 - \exp \left\{ - \sum_{u'}^u c_t \exp(\theta_{jkt}) \right\}, \quad j = 1, \dots, 41, \quad k \in \text{Trt}_j, \quad [u', u] \in \text{Tim}_j, \\ \theta_{jku} &= \mu_j + (v_u + \varphi_{ju})I(u \geq 2) + X_{jk}\delta_{jbk}, \\ \varphi_{ju} &\sim N(0, \tau_\varphi^2), \\ \delta_{jbk} &\sim N(d_k - d_b, \sigma_{bk}^2), \\ \text{prior : } \mu_j, v_u, d_k &\sim N(0, 10^3), \quad \tau_\varphi \sim U(0, 10), \quad \sigma_{bk} \sim \text{half-t}(0, \tau, \gamma), \end{aligned} \quad (4.4)$$

where  $p_{jku'u}$  is the healing probability for a patient in trial  $j$ , under treatment  $k$ , during the successive time segments in  $[u', u]$ ;  $\theta_{jku}$  the corresponding log hazard in time period  $u$ ;  $\text{Trt}_j$  is the set of treatments

involved in trial  $j$ ; and  $\text{Tim}_j$  the set of observed time periods for trial  $j$ . Comments on the weakly informative specification for the hyperparameters can be referred to Warn *and others* (2002), Spiegelhalter *and others* (2004), and Gelman (2006).

An apparent feature of the treatment regime in GERD data is that both therapies H<sub>2</sub>RA and PPI were used in single and double doses, which are regarded as different treatments. It seems likely that the effects of 2 treatments based on the same therapy but with different doses are more likely to have higher correlation than that based on different therapies. Furthermore, the placebo effects may have low PC with other treatments. For other correlations, we assume symmetry about 0.5. Based on these brief considerations, one possible choice for the prior correlations may be summarized in a matrix form, denoted by  $\mathbf{R}_0$ :

$$\begin{pmatrix} 1 & \text{Beta}(1.5, 3.5) & \text{Beta}(1.5, 3.5) & \text{Beta}(1.5, 3.5) & \text{Beta}(1.5, 3.5) & \text{Beta}(1.5, 3.5) \\ & 1 & \text{Beta}(2.5, 2.5) & \text{Beta}(2.5, 2.5) & \text{Beta}(2.5, 2.5) & \text{Beta}(2.5, 2.5) \\ & & 1 & \text{Beta}(8, 2) & \text{Beta}(2.5, 2.5) & \text{Beta}(2.5, 2.5) \\ & & & 1 & \text{Beta}(2.5, 2.5) & \text{Beta}(2.5, 2.5) \\ & & & & 1 & \text{Beta}(8, 2) \\ & & & & & 1 \end{pmatrix}.$$

The equivalent data used in the EPS method are just the parameters in these beta distributions. We then embed the “likelihood”  $\prod \rho_{bk}^{\alpha_{bk}} (1 - \rho_{bk})^{\beta_{bk}}$  into the model and call on the Cholesky parameterization (3.5) and the ancillary representation (3.3). The prior distributions for  $\tau_k^2$  may be specified as the IG distribution  $\text{IG}(\alpha, \beta)$ , where  $\alpha, \beta$  are given a weakly informative prior distribution  $\text{Exp}(0.01)$ .

For the GERD data, incorporating prior information  $\mathbf{R}_0$  into the PC model leads to a close model fit (with  $\text{DIC} = 257.4$  compared with  $\text{DIC} = 256.3$  for PC model without  $\mathbf{R}_0$ ) and similar posterior estimates for the population treatment contrasts but slightly narrower CIs. The impact of incorporating the prior correlation matrix  $\mathbf{R}_0$  is seen in the Figure 3 of posterior estimates of variances and their 95% CIs

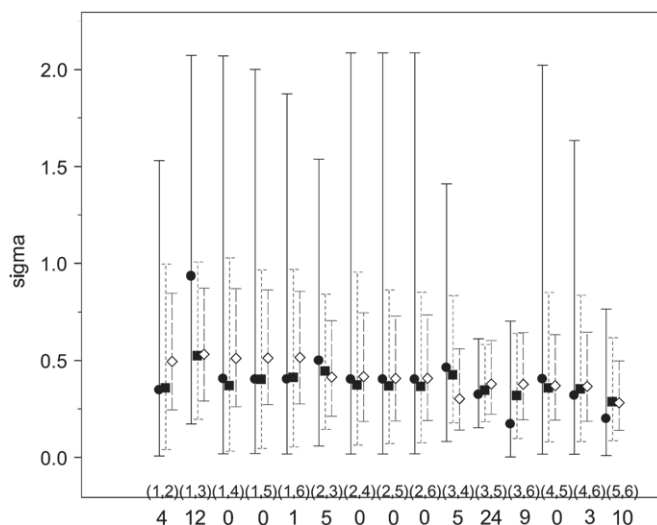


Fig. 3. Posterior medians of  $\{\sigma_{bk}\}$  with 95% CIs: (●) under half-t model, (■) under PC model prior  $\phi \sim U(0, \pi/2)$ , and (◇) incorporating prior information  $\mathbf{R}_0$ . The figures below case indices are the numbers of studies that make corresponding comparisons.

under 3 respective variance assumptions: (i) with the exchangeable half-t prior distribution for  $\sigma_{bk}$  (named EX model); (ii) with a PC structure, that is, the PC model with weakly informative priors  $U(0, \pi/2)$  for the angle parameters; (iii) the PC model incorporating the prior correlation information summarized in  $\mathbf{R}_0$ . Without the constraints of triangle inequalities, the EX model leads to much wider CIs for most variances than the other 2 models and also an outlier at the (1, 3) comparison. The narrowest CIs for all  $\sigma_{bk}$  occur in the third model, reflecting the effects of incorporating the prior information. Thus, incorporating variance structure stabilizes the posterior estimates of variance, particularly where there is little direct information in the likelihood, while adding expert knowledge about treatment similarity can further reduce prediction uncertainty in a coherent way.

## 5. SUMMARY

The variance structure inherent in MTC rests on the fundamental requirement that evidence from a variety of sources should be consistent, which logically leads to the consistency equations (2.2) and the triangle inequalities (3.1). We call the requirement of evidence consistency “fundamental” because, without it, we cannot perform a coherent synthesis of information on all the treatment contrasts. The crucial technical difficulty arises when we attempt to embed the triangle inequalities into MTC modeling. We circumvented this by adopting the ancillary representation (3.3), taking the separation strategy (3.4), and employing the spherical coordinates for Cholesky’s decomposition (3.5). We thus have the GC- and PC models that provide a way to specify weakly informative prior distributions for multiple heterogeneity parameters without violation of the triangle inequalities. When there is some prior information about the correlations between treatment arms, the EPS method allows the use of equivalent data for shipping the prior information into the analytic procedure. Finally, all the computations can be quickly done by MCMC simulations using WinBUGS or R.

## ACKNOWLEDGMENTS

This research was supported by Medical Research Council funding to the MRC Health Services Research Collaboration. The authors are thankful to John Copas, Simon Thomson, and other colleagues for their constructive comments when the early version of the paper was represented in *RSS Methods in Meta-analysis Meeting* held on 4 June 2007, in London, and to the editor Peter Diggle and anonymous referees for their insightful comments and suggestions. *Conflict of Interest*: None declared.

## FUNDING

Medical Research Council funding to the MRC Health Services Research Collaboration.

## SUPPLEMENTARY MATERIAL

Supplementary material is available at <http://biostatistics.oxfordjournals.org>.

## REFERENCES

- ADES, A. E., LU, G. AND HIGGINS, J. P. T. (2005). The interpretation of random effects meta-analysis in decision models. *Medical Decision Making* **25**, 646–654.
- ANDERSON, T. W., OLKIN, I. AND UNDERHILL, L. G. (1987). Generation of random orthogonal matrices. *SIAM Journal on Scientific and Statistical Computing* **8**, 625–629.

- BARNARD, J., McCULLOCH, R. AND MENG, X. L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica* **10**, 1281–1311.
- BEDFORD, T. AND COOKE, R. (2001). *Probability Risk Analysis: Foundations and Methods*. Cambridge: Cambridge University Press.
- CALDWELL, D. M., ADES, A. E. AND HIGGINS, J. P. T. (2005). Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *British Medical Journal* **331**, 897–900.
- DANIELS, M. J. AND KASS, R. E. (1999). Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association* **94**, 1254–1263.
- DEMPSTER, A. P. (1972). Covariance selection. *Biometrics* **28**, 157–175.
- EATON, M. L. AND OLKIN, I. (1987). Best equivariant estimators of a Cholesky decomposition. *Annals of Statistics* **15**, 1639–1665.
- FIGLIORE, M. C., BAILEY, W. C., COHEN, S. J., DORFMAN, S. F., GOLDSTEIN, M. G., GRITZ, E. R., HEYMAN, R. B., HOLBROOK, J., JAEN, C. R., KOTTKE, T. E. and others (1996). *Smoking Cessation. Clinical Practice Guideline No. 18*. Rockville, MD: Agency for Health Care Policy and Research, U.S. Department of Health and Human Services. AHCPR Publication No. 96-0692.
- GELMAN, A. (2006). Prior distributions for variance parameters in hierarchical Models. *Bayesian Analysis* **1**, 515–533.
- GELMAN, A. (2007). *Running WinBugs and Open Bugs from R*. <http://www.stat.columbia.edu/~gelman/bugsR/runningbugs.html>.
- GOEREE, R., O'BRIEN, B., HUNT, R., BLACKHOUSE, G., WILLAN, A. AND WATSON, J. (1999). Economic evaluation of long term management strategies for erosive oesophagitis. *Pharmacoeconomics* **16**, 679–697.
- HASSELBLAD, V. (1998). Meta-analysis of multi-treatment studies. *Medical Decision Making* **18**, 37–43.
- HIGGINS, J. P. T. AND WHITEHEAD, A. (1996). Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* **15**, 2733–2749.
- LEONARD, T. AND HSU, J. S. (1992). Bayesian inference for a covariance matrix. *Annals of Statistics* **20**, 1669–1696.
- LIU, C. (1993). Bartlett's decomposition of the posterior distribution of the covariance for normal monotone ignorable missing data. *Journal of Multivariate Analysis* **46**, 497–512.
- LU, G. AND ADES, A. E. (2004). Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine* **23**, 3105–3124.
- LU, G. AND ADES, A. E. (2006). Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association* **101**, 447–459.
- LU, G., ADES, A. E., SUTTON, A. J., COPPER, N. J., BRIGGS, A. H. AND CALDWELL, D. M. (2007). Meta-analysis of mixed treatment comparisons at multiple follow-up times. *Statistics in Medicine* **26**, 3681–3699.
- LUMLEY, T. (2002). Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine* **21**, 2313–2324.
- O'HAGAN, A., BUCK, C. E., DANESHKHAH, A., EISER, J. R., GARTHWAITE, P. H., JENKINSON, D. J., OAKLEY, J. E. AND RAKOW, T. (2006). *Uncertain Judgments: Eliciting Experts' Probabilities*. Chichester, UK: Wiley.
- PINHEIRO, J. C. AND BATES, D. M. (1996). Unconstrained parameterisations for variance-covariance matrices. *Statistics and Computing* **6**, 289–296.
- SMITH, T. C., SPIEGELHALTER, D. J. AND THOMAS, A. (1995). Bayesian approaches to random-effects meta-analysis: a comparative study. *Statistics in Medicine* **14**, 2685–2699.
- SPIEGELHALTER, D. J., ABRAMS, K. R. AND MYLES, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester, Chichester, UK: Wiley.

- SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. AND VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B* **64**, 1–34.
- SPIEGELHALTER, D. J., THOMAS, A., BEST, N. AND LUNN, D. (2001). *WinBUGS User Manual*. Version 1.4. Cambridge: MRC Biostatistics Unit.
- SUN, D. AND SUN, X. (2005). Estimation of the multivariate normal precision and covariance matrices in a star-shape model. *Annals of the Institute of Statistical Mathematics* **57**, 455–484.
- SUTTON, A. J. AND HIGGINS, J. P. T. (2008). Recent developments in meta-analysis. *Statistics in Medicine* **27**, 625–650.
- THOMPSON, S. G., SMITH, T. C. AND SHARP, S. J. (1997). Investigating underlying risk as a source of heterogeneity in meta-analysis. *Statistics in Medicine* **16**, 2741–2758.
- WARN, D. E., THOMPSON, S. G. AND SPIEGELHALTER, D. J. (2002). Bayesian random effects meta-analysis of trials with binary outcomes: method for absolute risk difference and relative risk scales. *Statistics in Medicine* **21**, 1601–1623.
- WINKLER, R. L. (1967). The assessment of prior distributions in Bayesian analysis. *Journal of the American Statistical Association* **67**, 776–800.
- YANG, R. AND BERGER, J. O. (1994). Estimation of a covariance matrix using the reference prior. *Annals of Statistics* **22**, 1195–1211.

[Received January 10, 2008; first revision October 20, 2008; second revision April 8, 2009; third revision May 26, 2009; fourth revision June 22, 2009; accepted for publication July 24, 2009]