HHS Public Access

Author manuscript

J Am Stat Assoc. Author manuscript; available in PMC 2016 June 01.

Published in final edited form as:

J Am Stat Assoc. 2015 June ; 110(510): 528–544. doi:10.1080/01621459.2015.1006065.

Bayesian Inference for Multivariate Meta-regression with a Partially Observed Within-Study Sample Covariance Matrix

Hui Yao[a], **Sungduk Kim**[b], **Ming-Hui Chen**[c], **Joseph G. Ibrahim**[d], **Arvind K. Shah**[e], and **Jianxin Lin**[e]

[a]Financial Services Office, Ernst & Young, New York, NY, USA

[b]Division of Intramural Population Health Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, NIH, Rockville, MD, USA

[c]Department of Statistics, University of Connecticut, Storrs, CT, USA

[d]Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA

[e]Clinical Biostatistics, Merck Research Laboratories, Rahway, NJ, USA

Summary

Multivariate meta-regression models are commonly used in settings where the response variable is naturally multi-dimensional. Such settings are common in cardiovascular and diabetes studies where the goal is to study cholesterol levels once a certain medication is given. In this setting, the natural multivariate endpoint is Low Density Lipoprotein Cholesterol (LDL-C), High Density Lipoprotein Cholesterol (HDL-C), and Triglycerides (TG) (LDL-C, HDL-C, TG). In this paper, we examine study level (aggregate) multivariate meta-data from 26 Merck sponsored double-blind, randomized, active or placebo-controlled clinical trials on adult patients with primary hypercholesterolemia. Our goal is to develop a methodology for carrying out Bayesian inference for multivariate meta-regression models with study level data when the within-study sample covariance matrix S for the multivariate response data is partially observed. Specifically, the proposed methodology is based on postulating a multivariate random effects regression model with an unknown within-study covariance matrix Σ in which we treat the within-study sample correlations as missing data, the standard deviations of the within-study sample covariance matrix S are assumed observed, and given Σ, S follows a Wishart distribution. Thus, we treat the off-diagonal elements of S as missing data, and these missing elements are sampled from the appropriate full conditional distribution in a Markov chain Monte Carlo (MCMC) sampling scheme via a novel transformation based on partial correlations. We further propose several structures (models) for Σ, which allow for borrowing strength across different treatment arms and trials. The proposed methodology is assessed using simulated as well as real data, and the results are shown to be quite promising.

Keywords

Aggregate covariates; Heterogeneity; Normal regression models; variable selection; Multiple trials; Random effects

1 Introduction

Multivariate responses in meta-regression models are commonly used when there is more than one endpoint of interest across studies, such as multiple outcomes, multiple time points, multiple treatments, and so forth. Such multiple outcomes are typically correlated and the sample correlations among these multiple outcomes are not typically reported in most published studies. One of the major challenges in multivariate meta-analysis with study level data is to conduct inference on the parameters in the meta-regression model when the within-study sample covariance (or correlation) matrix of these multiple outcomes is only partially observed. With individual patient data (IPD), the within-study sample covariance matrix is readily available since there is information in the data for computing this sample covariance matrix. However, for study level meta-data, inference in multivariate meta-regression with a partially observed within-study sample covariance matrix is a long-standing and difficult problem, with no gold standard solutions.

There has been some literature addressing this problem both from a Bayesian and frequentist perspective. When only the diagonal elements of the within-study sample covariance matrix S are observed, one simple remedy is to impute the missing sample correlations over the entire range of values (i.e., from -1 to 1) and then assess whether the conclusions depend on the correlations that are imputed. This type of analysis has been used in a multivariate meta-analysis of 44 trials which evaluated the effectiveness of injectable gold, auranofin and placebo on three treatment outcomes (Berkey et al., 1996). Nam et al. (2003) propose and evaluate three Bayesian multivariate meta-analysis models. In the case of bivariate outcomes, they assume a uniform distribution on $(-1, 1)$ for each within-study correlation. For a bivariate random-effects meta-analysis, Riley et al. (2008) propose a model which does not require knowing the within-study sample correlations. Their model includes only one overall correlation parameter, which can be considered a hybrid measure of the within-study and between-study correlations. Unless the overall correlation is very close to 1 or -1, this alternative model has been shown to produce appropriate pooled estimates with little bias. Wei and Higgins (2013a) examine a multivariate random effects meta-regression model from a frequentist perspective, where they estimate the within-study covariance matrix of the mean difference in the treatment effects and odds ratios assuming the within-study correlations are known. Wei and Higgins (2013b) discuss Bayesian multivariate meta-analysis with multiple outcomes with a known within-study covariance matrix, where they decompose the between-study covariance matrix into a product of variances and correlations as in Barnard et al. (2000), carry out a Cholesky decomposition of the between-study correlation matrix, and specify uniform priors on the Cholesky elements while at the same time ensuring positive definiteness. Ma and Mazumdar (2011) examine robust methods based on U-statistics for a multivariate meta-analysis random effects model assuming that the within-study sample covariance matrix is known. Hamza et al. (2009) examine multivariate random effects meta-analysis models with applications to diagnostic tests, where again, the within-study covariance matrix is assumed known. Hedges et al. (2010) provide a robust estimator of the covariance matrix of the meta-regression coefficients in the setting of clusters of internally correlated estimates. They only consider univariate aggregate responses and then assume that these aggregate responses are correlated within the same

cluster. Their paper does not examine the multivariate meta-regression setting nor does it examine inference for the within-study covariance matrix based on aggregate data. Their methodology and data structure are very different than the setting considered in this paper. Jackson et al. (2011) and Riley (2009) give very nice overviews and detailed reviews of multivariate meta-analysis methods and provide some practical guidance and recommendations as to how to specify and/or estimate the within-study covariance matrix. Both of these papers also have an exhaustive reference list on this topic.

In this paper, we present a fundamentally different approach than the aforementioned literature. Our primary goal is to carry out inferences on the parameters in multivariate meta-regression models with study level data in which the within-study sample covariance matrix S is only partially observed. Towards this goal, we propose a Bayesian multivariate random effects meta-regression model in which S follows a Wishart distribution, where the parameter of the Wishart distribution is the unknown within-study covariance matrix Σ. We aim to recover the unobserved elements of the within-study sample covariance matrix by modeling Σ and borrowing strength from different treatment arms across different trials. Since only the diagonal elements of S are observed and reported in most published studies, we derive the induced distribution of the off-diagonal elements of S via a decomposition of S similar to that of Barnard et al. (2000). We then devise a novel Markov chain Monte Carlo (MCMC) sampling algorithm for the sample correlation matrix R. Our approach has a flavor similar to that discussed in Riley (2009) but is really quite different, since we do not specify values of the sample correlation matrix R nor do we specify prior distributions on functions of the elements of R. In our approach, the sampling distribution for R is induced from the Wishart distribution that is imposed on S at the outset. By treating the off-diagonal elements of S as missing data, we can write the complete data likelihood function of the multivariate responses and S given Σ, specify the appropriate priors for Σ, and then carry out full Bayesian inference via MCMC methods. Our approach is analogous and similar in flavor to the one in the context of missing covariates in regression models as discussed in Ibrahim et al. (2005). Specifically, S is viewed as the partially observed "covariate data" and therefore is parametrically modeled, and inference on S is then carried out through the parameters (Σ) of the distribution of S.

The rest of this paper is organized as follows. In Section 2, we give a description of the cholesterol meta-data with three primary aggregate outcome variables from the 26 clinical trials. In Section 3, we give the full development of the multivariate meta-regression random effects model, the induced conditional distribution of the sample correlation matrix R given the observed diagonal elements of the within-study sample covariance matrix, and the complete data likelihood. In Section 4, we present a novel MCMC sampling algorithm based on sampling the partial correlations from R. In Section 5, we examine specific types of structures and models for the unknown within-study covariance matrix Σ that may be used in practice, and Section 6 gives the general Bayesian computational development and goodness-of-fit criterion for model comparisons. Section 7 presents an analysis of the cholesterol meta-data discussed in Section 2. In Section 8, we carry out two detailed simulation studies. We conclude the article with some discussion in Section 9.

2 Cholesterol Data

Millions of Americans are struggling with high cholesterol which is well known to contribute to heart disease and other cardiovascular disease. A great deal of effort has been put forth in clinical trials studying cholesterol lowering drugs. Endpoints in such trials typically focus on one or more of three primary endpoints, these being Low Density Lipoprotein Cholesterol (LDL-C), High Density Lipoprotein Cholesterol (HDL-C), and Triglycerides (TG) (LDL-C, HDL-C, TG). The multivariate aggregate meta-data come from 26 Merck sponsored double-blind, randomized, active or placebo-controlled clinical trials on adult patients with primary hypercholesterolemia. The primary goal of these clinical trials was to evaluate the effects of Ezetimibe (EZE) on LDL-C, HDL-C, and TG (which works in the digestive tract) in combination with statin (which works in the liver) in comparison to statin alone on treatment-naïve patients at baseline (on a first line therapy) and those continuing on statins at baseline (on a second line therapy). The citations of primary published papers in clinical journals for the 26 trials considered in this paper can be found in Leiter et al. (2011) and Chen et al. (2012). These trials were conducted between November 1999 to October 2008 and study durations ranged from 4 weeks to 24 weeks. Some trials had longer durations with titration of doses but only the data prior to the first titration were used in the analyses. The entry criteria for the patients in each of these studies are given in Chen et al. (2012).

The primary endpoints in these trials are the mean percent changes in LDL-C, HDL-C, and TG from their respective baseline values, denoted by (LDL-C, HDL-C, TG). The aggregate covariates considered in our analysis include treatment (trt) (“statin” or “statin+Ezetimibe”) on the first line therapy or on the second line therapy, baseline LDL-C (bl_ldlc), baseline HDL-C (bl_hdlc), and baseline TG (bl_tg), age in years, white (%), male (%), Diabetes Mellitus (DM) (%), and Duration in weeks (Dur). The multivariate aggregate meta-data including the aggregate covariates are given in Appendix A of the supplementary document. Figures 1 and 2 show the forest plots of the aggregate meta-data for three primary outcome variables for these 26 studies. In these two figures, each line corresponds to percent change in each of LDL-C, HDL-C, and TG from baseline ± one sample standard deviation. We note that the reported means were model-based means. The lines based on the mean ± one standard deviation shown in Figures 1 and 2 are much longer than the lines constructed based on 95% confidence intervals shown in the forest plot for LDL-C in Chen et al. (2012). We further note the sample correlations of the three primary aggregate outcome variables (LDL-C, HDL-C, TG) were not reported in published papers in the clinical journals for the 26 trials (Leiter et al., 2011). Thus, the aggregate meta-data from these 26 trials provide a great motivation for developing a new methodology using Bayesian multivariate random effects meta-regression models in the subsequent sections.

3 Methods for Multivariate Meta-regression and Multi-Dimensional Random Effects

3.1 The Multivariate Meta-regression Model

Consider K randomized trials, where each trial has T treatment arms. The sample size of the t^{th} treatment arm within the k^{th} trial is n_{tk} for $t = 1, \ldots, T$ and $k = 1, \ldots, K$. Let $\boldsymbol{y}_{itk} = (y_{itk1}, \ldots, y_{itkJ})'$ be the J-dimensional response of the i^{th} patient in the t^{th} treatment arm within the k^{th} study. Also let $\boldsymbol{x}_{tkj}$ denote a p_j-dimensional vector of treatment-within-trial level covariates for the j^{th} response corresponding to the fixed effects for the t^{th} treatment arm. We further let $\boldsymbol{z}_{tkj}$ denote a q_j-dimensional vector of treatment-within-trial level covariates corresponding to the random effects for the t^{th} treatment arm in the k^{th} trial.

The multivariate meta-regression model for $\boldsymbol{y}_{itk}$ assumes

$$\boldsymbol{y}_{itk}=\boldsymbol{\mu}_{tk}+\boldsymbol{\varepsilon}_{itk},\ \ \boldsymbol{\varepsilon}_{itk}=(\in_{itk1},\ldots,\in_{itkJ})'\sim N\left(\boldsymbol{\mu}_{tk},\sum\nolimits_{tk}\right),i=1,\ldots,n_{tk}, \quad (3.1)$$

where Σ_{tk} is the $J \times J$ covariance matrix, $\boldsymbol{\mu}_{tk} = (\mu_{tk1}, \ldots, \mu_{tkJ})'$, and

$$\mu_{tkj}=\boldsymbol{x}_{tkj}^{'}\boldsymbol{\beta}_j+\boldsymbol{z}_{tkj}^{'}\gamma_{kj},\ \ j=1,\ldots,J, \quad (3.2)$$

for $t = 1, \ldots, T$ and $k = 1, \ldots, K$. In (3.1), we assume that the $\boldsymbol{\varepsilon}_{itk}$'s are independent. In (3.2), $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jpj})'$ is the vector of fixed effects regression coefficients corresponding to the p_j covariates, and $\boldsymbol{\gamma}_{kj} = (\gamma_{kj1}, \ldots, \gamma_{kjqj})'$ represents the vector of q_j-dimensional random effects for the j^{th} response for $j = 1, \ldots, J$, $t = 1, \ldots, T$, and $k = 1, \ldots, K$. We further assume that

$$\boldsymbol{\gamma}_{kj}\sim N\left(\mathbf{0},\Omega_j\right), \quad (3.3)$$

where Ω_j is a $q_j \times q_j$ covariance matrix. The random effects $\boldsymbol{\gamma}_{kj}$'s, which are assumed to be independent of the $\boldsymbol{\varepsilon}_{itk}$, capture heterogeneity across the K trials for the j^{th} response.

Letting $\overline{y}_{\cdot tkj}=\frac{1}{n_{tk}}\sum_{i=1}^{n_{tk}} y_{itkj}$ for $j = 1, \ldots, J$, the J-dimensional sample mean and $J \times J$ sample covariance matrix are

$$\overline{\boldsymbol{y}}_{\cdot tk}=(\overline{y}_{\cdot tk1},\ldots,\overline{y}_{\cdot tkJ})' \text{ and } S_{tk}=\frac{1}{n_{tk}-1}\sum_{i=1}^{n_{tk}}(\boldsymbol{y}_{itk}-\boldsymbol{y}_{\cdot tk})(\boldsymbol{y}_{itk}-\boldsymbol{y}_{\cdot tk})' \quad (3.4)$$

for $t = 1, \ldots, T$ and $k = 1, \ldots, K$. Then, we have the following result.

Result 3.1—*The sample mean $\bar{\boldsymbol{y}}_{\cdot tk}$ and sample covariance S_{tk} are joint complete sufficient statistics for $(\boldsymbol{\mu}_{tk}, \Sigma_{tk})$. In addition, $\bar{\boldsymbol{y}}_{\cdot tk}$ and S_{tk} are independent. Consequently, we have*

$$\bar{y}_{\cdot tkj} = \mu_{tkj} + \bar{\varepsilon}_{\cdot tkj} = \boldsymbol{x}'_{tkj}\boldsymbol{\beta}_j + \boldsymbol{z}'_{tkj}\boldsymbol{\gamma}_{kj} + \bar{\varepsilon}_{\cdot tkj}, \quad (3.5)$$

where $\bar{\in}_{\cdot tkj} = \frac{1}{n_{tk}} \sum_{i=1}^{n_{tk}} \in_{itkj}$, *j* = 1, … *J and*

$$\bar{\boldsymbol{\varepsilon}}_{\cdot tk} = (\bar{\in}_{\cdot tk1}, \ldots, \bar{\in}_{\cdot tkJ})' \sim N\left(\mathbf{0}, \frac{1}{n_{tk}} \sum\nolimits_{tk}\right), \quad (3.6)$$

and

$$(n_{tk} - 1)\, S_{tk} \sim W_{n_{tk}-1}\left(\sum\nolimits_{tk}\right), \quad (3.7)$$

where $W_{n_{tk}-1}(\Sigma_{tk})$ *denotes the Wishart distribution with* $n_{tk} - 1$ *degrees of freedom and positive definite* $J \times J$ *scale matrix* Σ_{tk}.

We briefly discuss the proof of Result 3.1 using Basu's Theorem (Lehmann and Casella, 1998). The joint pdf of $\boldsymbol{y}_{itk}$ for $i = 1, \ldots, n_{tk}$ can be written as,

$$\begin{aligned} f(\boldsymbol{y}_{1tk}, \ldots, \boldsymbol{y}_{n_{tk}tk} | \boldsymbol{\mu}_{tk}, \sum\nolimits_{tk}) &= [(2\pi)^J |\sum\nolimits_{tk}|]^{-n_{tk}/2} \exp\{-\tfrac{1}{2} \sum_{i=1}^{n_{tk}} [(\boldsymbol{y}_{itk} - \boldsymbol{\mu}_{tk})' \sum\nolimits_{tk}^{-1} (\boldsymbol{y}_{itk} - \boldsymbol{\mu}_{tk})]\} \\ &= [(2\pi)^J |\sum\nolimits_{tk}|]^{-n_{tk}/2} \exp\left\{-\tfrac{1}{2} \sum_{i=1}^{n_{tk}} [\boldsymbol{y}'_{itk} \sum\nolimits_{tk}^{-1} \boldsymbol{y}_{itk} - 2\boldsymbol{\mu}'_{tk} \sum\nolimits_{tk}^{-1} \boldsymbol{y}_{itk} + \boldsymbol{\mu}'_{tk} \sum\nolimits_{tk}^{-1} \boldsymbol{\mu}_{tk}]\right\}. \end{aligned}$$

Due to the properties of the exponential family, $\left(\sum_{i=1}^{n_{tk}} \boldsymbol{y}_{itk}, \sum_{i=1}^{n_{tk}} \boldsymbol{y}_{itk} \boldsymbol{y}'_{itk}\right)$ are the joint complete sufficient statistics for $\boldsymbol{\mu}_{tk}$ and Σ_{tk}. Since $\bar{\boldsymbol{y}}_{\cdot tk}$ and S_{tk} are one-to-one functions of $\sum_{i=1}^{n_{tk}} \boldsymbol{y}_{itk}$ and $\sum_{i=1}^{n_{tk}} \boldsymbol{y}_{itk} \boldsymbol{y}'_{itk}$, they are jointly complete and sufficient. Now for any fixed Σ_{tk}, $\bar{\boldsymbol{y}}_{\cdot tk}$ is complete sufficient for $\boldsymbol{\mu}_{tk}$ and S_{tk} is ancillary for $\boldsymbol{\mu}_{tk}$. Then by Basu's Theorem, $\bar{\boldsymbol{y}}_{\cdot tk}$ and S_{tk} are independent. The results given in (3.6) and (3.7) immediately follow from (3.1) and the independence assumption of the $\boldsymbol{\varepsilon}_{itk}$'s.

The classical meta-regression model (e.g., Whitehead, 2002) often assumes

$$\bar{\boldsymbol{\varepsilon}}_{\cdot tk} \sim N\left(\mathbf{0}, \frac{1}{n_{tk}} S_{tk}\right). \quad (3.8)$$

However, only $\bar{y}_{\cdot tk}$ and the diagonal elements of S_{tk} are available in most published articles in the literature. Therefore, the multivariate meta-regression model defined by (3.3), (3.5), and (3.8) is not identifiable due to the missing values of the off-diagonal elements in S_{tk}. In the remainder of this paper, we assume that only the treatment-within-trial level response $\bar{y}_{\cdot tk}$ and the diagonal elements of S_{tk} are available while the individual patient level responses, $\boldsymbol{y}_{itk}$, as well as the off-diagonal elements of S_{tk} are not available.

Equations (3.6) and (3.7) can be viewed as a natural extension of Yao et al. (2011) for the univariate meta-regression model. We use the Wishart distribution with parameter Σ_{tk} in (3.7) to help "recover" the missing off-diagonal elements of S_{tk} and to facilitate the joint estimation of $\boldsymbol{\beta}_j$, Ω_j, and Σ_{tk}. In addition, using (3.7), the probability density function (pdf) of S_{tk} given Σ_{tk} can be written as

$$f\left(S_{tk}|\sum_{tk}\right) = (n_{tk}-1)^{\frac{J(J+1)}{2}} \left(2^{\frac{(n_{tk}-1)J}{2}} \pi^{\frac{J(J-1)}{4}} \prod_{j=1}^{J} \left(\frac{n_{tk}-j}{2}\right)\right)^{-1} \times |\sum_{tk}|^{-\frac{n_{tk}-1}{2}} |(n_{tk}-1)\, S_{tk}|^{\frac{n_{tk}-J-2}{2}} \exp\{-\tfrac{1}{2} tr((n_{tk}-1)\sum_{tk}^{-1} S_{tk})\}. \quad (3.9)$$

Let V_{tk} = diag(S_{tk11}, …, S_{tkJJ}) be the diagonal matrix that contains the diagonal elements of S_{tk}. Then, the sample correlation matrix is given by

$$R_{tk} = V_{tk}^{-\frac{1}{2}} S_{tk} V_{tk}^{-\frac{1}{2}}. \quad (3.10)$$

Each diagonal element of R_{tk} is equal to 1, while each off-diagonal element lies between −1 and 1. Prom the distribution of S_{tk} in (3.7), the conditional distribution of R_{tk} can be written as

$$f\left(R_{tk}|V_{tk}, \sum_{tk}\right) \propto \left|\sum_{tk}\right|^{-\frac{n_{tk}-1}{2}} \left|V_{tk}^{\frac{1}{2}} R_{tk} V_{tk}^{\frac{1}{2}}\right|^{\frac{n_{tk}-J-2}{2}} \exp\{-\frac{1}{2} tr((n_{tk}-1)\sum_{tk}^{-1} V_{tk}^{\frac{1}{2}} R_{tk} V_{tk}^{\frac{1}{2}})\}. \quad (3.11)$$

Remark 3.1: In Result 3.1, we assume that μ_{tkj} depends only on the trial-level covariates. This assumption holds only when patients with the same characteristics are selected by a trial. But for most clinical trials, values of the individual-level covariates may vary within trial and treatment subgroups. This assumption can still be met for randomized trials in which patient-specific covariates are randomized. For this case, the regression coefficients capture the difference among trials, and μ_{tkj} is affected only by the difference in characteristics across trials. Therefore, Result 3.1 is applicable to meta-data from randomized trials.

Remark 3.2: The aggregate response, $\bar{\boldsymbol{y}}_{\cdot tk} = (\bar{y}_{\cdot tk1}, \ldots, \bar{y}_{\cdot tkJ})'$, and the diagonal matrix of the sample covariance S_{tk}, V_{tk} = diag(S_{tk11}, …, S_{tkJJ}), are observed for $t = 1, \ldots, T$ and $k = 1, \ldots, K$. In Result 3.1, (3.5) allows us to model Σ_{tk} and the distribution of the sample covariance S_{tk} in (3.6) yields the induced distribution (3.11) of the sample correlation matrix R_{tk} given Σ_{tk} and V_{tk}. We note here that $\bar{\boldsymbol{y}}_{\cdot tk}$ is a vector of multivariate aggregate responses, which does carry the information of the correlations among these multivariate responses. Once the covariance matrix Σ_{tk} is modeled appropriately, the information of the missing sample correlation matrix R_{tk} can be recovered via (3.6), especially when TK is large. By treating the off-diagonal elements of S as missing data, (3.5) essentially specifies the distribution of "response variable" while (3.6) induces the distribution of "missing covariates" in regression models as discussed in Ibrahim et al. (2005). However, our approach differs from the

standard missing covariates problem in the sense that (i) the missing sample correlation matrix, treating as "missing coavriates", does not directly enter the "response" model and (ii) the "response" model and the "missing covariates" model are connected via the unknown covariance matrix Σ_{tk}.

3.2 The Complete-data Likelihood Function

Due to the independence of $\bar{\boldsymbol{y}}_{\cdot tk}$ and S_{tk} shown in the last section, their joint density can be written product of marginal densities as

$$\begin{aligned} f\left(\overline{\boldsymbol{y}}_{\cdot tk}, S_{tk} | \boldsymbol{\beta}_j, \textstyle\sum_{tk}, \boldsymbol{\gamma}_{kj}\right) &= (2\pi)^{-\frac{J}{2}} \left|\frac{\sum_{tk}}{n_{tk}}\right|^{-\frac{1}{2}} \exp\{-\tfrac{n_{tk}}{2}\ (\overline{\boldsymbol{y}}_{\cdot tk} - \boldsymbol{\mu}_{tk})' \textstyle\sum_{tk}^{-1} (\overline{\boldsymbol{y}}_{\cdot tk} - \boldsymbol{\mu}_{tk})\} \\ &\times (n_{tk}-1)^{\frac{J(J+1)}{2}} \left\{ 2^{\frac{(n_{tk}-1)J}{2}} \pi^{\frac{J(J-1)}{4}} \prod_{j=1}^{J} \left(\tfrac{n_{tk}-j}{2}\right) \right\}^{-1} |\textstyle\sum_{tk}|^{-\frac{n_{tk}-1}{2}} \\ &\times |(n_{tk}-1)\, S_{tk}|^{\frac{n_{tk}-J-2}{2}} \exp\left[-\tfrac{1}{2} tr\{(n_{tk}-1)\textstyle\sum_{tk}^{-1} S_{tk}\}\right]. \end{aligned}$$

Let $D_c = \{(\bar{\boldsymbol{y}}_{\cdot tk}, S_{tk}, n_{tk}, \boldsymbol{x}_{tkj}, \boldsymbol{z}_{tkj}, \boldsymbol{\gamma}_{kj}), t = 1, \ldots, T, j = 1, \ldots, J, k = 1, \ldots, K\}$. The complete-data likelihood function for the model in (3.5) and (3.6) is thus given by

$$\begin{aligned} L_c(\boldsymbol{\beta}, \textstyle\sum^{*} | D_c) &\propto \prod_{t=1}^{T} \prod_{k=1}^{K} \left(\left|\frac{\sum_{tk}}{n_{tk}}\right|^{-\frac{1}{2}} \exp\{-\tfrac{n_{tk}}{2} (\overline{\boldsymbol{y}}_{\cdot tk} - \boldsymbol{\mu}_{tk})' \textstyle\sum_{tk}^{-1} (\overline{\boldsymbol{y}}_{\cdot tk} - \boldsymbol{\mu}_{tk})\} \right. \\ &\left. \times |\textstyle\sum_{tk}|^{-\frac{n_{tk}-1}{2}} |(n_{tk}-1) S_{tk}|^{\frac{n_{tk}-J-2}{2}} \exp\left[-\tfrac{1}{2} tr\{(n_{tk}-1)\textstyle\sum_{tk}^{-1} S_{tk}\}\right] \right), \end{aligned} \tag{3.12}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^{'}, \ldots, \boldsymbol{\beta}_J^{'})'$ and $\boldsymbol{\Sigma}^* = (\Sigma_{11}, \ldots, \Sigma_{T1}, \ldots, \Sigma_{1K}, \ldots, \Sigma_{TK})$.

4 Sampling the Correlation Matrix *R*

The conditional distribution of the sample correlations in (3.11) is critical in recovering the unobserved correlations for jointly modeling the multivariate aggregate responses. However, it is difficult to sample directly from the conditional density in (3.11) due to the constraint of positive definiteness for R_{tk}. To relax the requirement of positive definiteness, we develop a sampling algorithm for R_{tk} via partial correlations using the techniques of Daniels and Pourahmadi (2001) and Wang and Daniels (2013). For ease of presentation, we drop the index "*tk*" in R_{tk} V_{tk}, and Σ_{tk} in this section.

Let $R = (r_{ij})$ be a $J \times J$ positive definite correlation matrix. Then, R can be parameterized in terms of the correlations $r_{i,i+j}$ for $i = 1, \ldots, J-1$ and the partial correlations $r_{i,j|i+1,\ldots,j-1}$ for $j - i \geq 2$. To make the transformation of the correlations $r_{j,j+k}$ to the partial correlations $r_{j,j+k|j+1,\ldots,j+k-1}$, we define $R[j : j+k] = R[j : j+k, j : j+k]$ as the $(k+1) \times (k+1)$ sub-matrix of R, which takes elements from the j^{th} row to the $(j+k)^{th}$ row and the j^{th} column to the $(j+k)^{th}$ column. It can be partitioned as

$$R[j{:}j+k]=\begin{pmatrix} 1 & \boldsymbol{r}_1(j,k) & r_{j,j+k} \\ \boldsymbol{r}_1'(j,k) & R_2(j,k) & \boldsymbol{r}_3'(j,k) \\ r_{j+k,j} & \boldsymbol{r}_3(j,k) & 1 \end{pmatrix}, \quad (4.1)$$

where $\boldsymbol{r}_1(j,k) = (r_{j,j+1}, \ldots, r_{j,j+k-1})$, $\boldsymbol{r}_3(j,k) = (r_{j+k,j+1}, \ldots, r_{j+k,j+k-1})$, $R_2(j,k)$ contains the middle $k-1$ rows and columns of $R[j : j+k]$. The partial correlations can be written function of the marginal correlations,

$$r_{j,j+k|j+1,\ldots,j+k-1}=\frac{r_{j,j+k}-\boldsymbol{r}_1(j,k)R_2^{-1}(j,k)\boldsymbol{r}_3'(j,k)}{[1-\boldsymbol{r}_1(j,k)R_2^{-1}(j,k)\boldsymbol{r}_1'(j,k)]^{1/2}[1-\boldsymbol{r}_3(j,k)R_2^{-1}(j,k)\boldsymbol{r}_3'(j,k)]^{1/2}}, \quad (4.2)$$

for $2 \le k \le J-j$.

This also leads to a formula for $r_{j,j+k}$, given by

$$r_{j,j+k}=\boldsymbol{r}_1(j,k)R_2^{-1}(j,k)\boldsymbol{r}_3'(j,k)+r_{j,j+k|j+1,\ldots,j+k-1}A_{jk}, \quad (4.3)$$

where $A_{jk}=[1-\boldsymbol{r}_1(j,k)R_2^{-1}(j,k)\boldsymbol{r}_1'(j,k)]^{1/2}[1-\boldsymbol{r}_3(j,k)R_2^{-1}(j,k)\boldsymbol{r}_3'(j,k)]^{1/2}$.

One advantage of this reparameterization is that $r_{j,k|j+1,\ldots,j+k-1}$ can vary independently in $(-1, 1)$ while maintaining positive definiteness of R. Hence to generate a random correlation matrix, one may generate partial correlations independently and then transform back to $r_{j,j+k}$ for $2 \le k \le J-1$. Under Sylvester's criterion, a necessary and sufficient criterion for R to be positive definite is that all of the leading principal minors must be positive. In other words, all the following matrices have a positive determinant: (i) the upper left 1-by-1 corner of R; (ii) the upper left 2-by-2 corner of R; (iii) the upper left 3-by-3 corner of R; …; and (iv) R itself. Joe (2006) proposed that the determinant of R can be calculated as

$$\det(R)=\prod_{i=1}^{J-1}\left(1-r_{i,i+1}^2\right)\times\prod_{k=2}^{J-1}\prod_{j=1}^{J-k}\left(1-r_{j,j+k|j+1,\ldots,j+k-1}^2\right). \quad (4.4)$$

We can use this property to show that the determinants of R and all the leading principal minors are positive when $-1 < r_{i,i+1} < 1$ and $-1 < r_{j,j+k|j+1,\ldots,j+k-1} < 1$ for $i = 1, \ldots, J-1$, $k = 2, \ldots, J-1$, and $j = 1, \ldots, J-k$.

For purposes of illustration, we consider a 3–dimensional correlation matrix

$$R=\begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{pmatrix}.$$

We now show how to generate the correlation matrix R through the use of partial correlations where R has a conditional density specified in (3.11). It can be shown that the determinant of the Jacobian for the transformation of (r_{12}, r_{23}, r_{13}) to (r_{12}, r_{23}, $r_{13|2}$) is $[(1-r_{12}^2)(1-r_{23}^2)]^{-1/2}$. Writing the correlations in terms of their partial correlation counterparts, we have

$$R(r_{13|2}) = \begin{pmatrix} 1 & r_{12} & r_{13}(r_{13|2}) \\ r_{12} & 1 & r_{23} \\ r_{13}(r_{13|2}) & r_{23} & 1 \end{pmatrix},$$

where $r_{13}(r_{13|2}) = r_{12}r_{23} + r_{13|2}(1-r_{12}^2)^{1/2}(1-r_{23}^2)^{1/2}$. Thus, the density of the partial correlations $R(r_{13|2})$ can be written as

$$\begin{aligned} g*(R(\rho_{13|2})|V,\textstyle\sum) &\propto |\textstyle\sum|^{-\frac{n-1}{2}} |V^{\frac{1}{2}} R(\rho_{13|2}) V^{\frac{1}{2}}|^{\frac{n-d-2}{2}} \\ \times \exp\{-\tfrac{1}{2} tr((n-1) R(\rho_{13|2}) V^{\frac{1}{2}} \textstyle\sum^{-1} V^{\frac{1}{2}})\} &[(1-\rho_{12}^2)(1-\rho_{23}^2)]^{-1/2}. \end{aligned} \quad (4.5)$$

We further make the Fisher's z transformation on each partial correlation, leading to $z_{12} = \frac{1}{2}\log\frac{1+r_{12}}{1-r_{12}}$, $z_{23} = \frac{1}{2}\log\frac{1+r_{23}}{1-r_{23}}$, and $z_{13|2} = \frac{1}{2}\log\frac{1+r_{13|2}}{1-r_{13|2}}$. The Jacobian for this transformation is $\frac{4e^{2z_{12}}}{(e^{2z_{12}}+1)^2}\frac{4e^{2z_{23}}}{(e^{2z_{23}}+1)^2}\frac{4e^{2z_{13|2}}}{(e^{2z_{13|2}}+1)^2}$. Rewriting the correlation matrix again in terms of functions of the Fisher's $\boldsymbol{z}$ values and letting $\boldsymbol{z} = (z_{12}, z_{23}, z_{13|2})'$, we get

$$R(\boldsymbol{z}) = \begin{pmatrix} 1 & r_{12}(\boldsymbol{z}) & r_{13}(\boldsymbol{z}) \\ r_{12}(\boldsymbol{z}) & 1 & r_{23}(\boldsymbol{z}) \\ r_{13}(\boldsymbol{z}) & r_{23}(\boldsymbol{z}) & 1 \end{pmatrix},$$

where $r_{12}(\boldsymbol{z}) = \frac{e^{2z_{12}}-1}{e^{2z_{12}}+1}$, $r_{23}(\boldsymbol{z}) = \frac{e^{2z_{23}}-1}{e^{2z_{23}}+1}$, and $r_{13}(\boldsymbol{z}) = \frac{e^{2z_{12}}-1}{e^{2z_{12}}+1}\frac{e^{2z_{23}}-1}{e^{2z_{23}}+1} + \frac{e^{2z_{13|2}}-1}{e^{2z_{13|2}}+1}\left(1-\left\{\frac{e^{2z_{12}}-1}{e^{2z_{12}}+1}\right\}^2\right)^{1/2}\left(1-\left\{\frac{e^{2z_{23}}-1}{e^{2z_{23}}+1}\right\}^2\right)^{1/2}$. The density of $R(\boldsymbol{z})$ is therefore

$$\begin{aligned} g(\boldsymbol{z}|V,\textstyle\sum) \propto |\textstyle\sum|^{-\frac{n-1}{2}} |V^{\frac{1}{2}} R(\boldsymbol{z}) V^{\frac{1}{2}}|^{\frac{n-J-2}{2}} &\exp\{-\tfrac{1}{2} tr((n-1) R(\boldsymbol{z}) V^{\frac{1}{2}} \textstyle\sum^{-1} V^{\frac{1}{2}})\} \\ \times \left[1-\left(\tfrac{e^{2z_{12}}-1}{e^{2z_{12}}+1}\right)^2\right]^{-1/2} &\left[1-\left(\tfrac{e^{2z_{23}}-1}{e^{2z_{23}}+1}\right)^2\right]^{-1/2} \frac{4e^{2z_{12}}}{(e^{2z_{12}}+1)^2}\frac{4e^{2z_{23}}}{(e^{2z_{23}}+1)^2}\frac{4e^{2z_{13|2}}}{(e^{2z_{13|2}}+1)^2}. \end{aligned} \quad (4.6)$$

One may sample directly from (4.6) using the Adaptive Rejection Metropolis Sampling (ARMS) algorithm. Alternatively, the Fisher's $\boldsymbol{z}$'s can be generated by a localized Metropolis algorithm (Chen et al., 2000), which entails the following steps:

Step 1 Let $\boldsymbol{z}^{(m-1)}$ be the current value.

Step 2 Draw $\boldsymbol{z}^*$ from $N(\hat{\boldsymbol{z}}, \Sigma_z)$.

Step 3 Compute

$$a = \min\left\{\frac{g(\boldsymbol{z}^*|V,\sum)\phi\{(\boldsymbol{z}^{(m-1)}-\hat{\boldsymbol{z}})'\sum_z^{-1}(\boldsymbol{z}^{(m-1)}-\hat{\boldsymbol{z}})\}}{g(\boldsymbol{z}^{(m-1)}|V,\sum)\phi\{(\boldsymbol{z}^*-\hat{\boldsymbol{z}})'\sum_z^{-1}(\boldsymbol{z}^*-\hat{\boldsymbol{z}})\}}, 1\right\}$$

where φ is the standard normal density function.

Step 4 Draw u from $U(0, 1)$. Set $\boldsymbol{z}^{(m)} = \boldsymbol{z}^*$ if $u \le a$ and $\boldsymbol{z}^{(m)} = \boldsymbol{z}^{(m-1)}$ if $u > a$.

Here, $N(\hat{\boldsymbol{z}}, \Sigma_z)$ is the normal proposal density, where $\hat{\boldsymbol{z}}$ is the maximizer of the logarithm of $g(\boldsymbol{z}|V, \Sigma)$, and Σ_z is minus of the inverse of the second derivative of log $g(\boldsymbol{z}|V, \Sigma)$ evaluated at $\boldsymbol{z} = \boldsymbol{z}^*$. That is,

$$\sum_z^{-1} = \frac{\partial^2 \log g(\boldsymbol{z}|V,\sum)}{\partial \boldsymbol{z}\partial \boldsymbol{z}'}\Big|_{\boldsymbol{z}=\boldsymbol{z}^*}.$$

5 Models for Σ_{tk}

As we will show in Section 8.1, the missing off-diagonal elements of S_{tk} can well be recovered once Σ_{tk} is correctly specified. Unlike S_{tk}, Σ_{tk} can be taken to have certain structures (i.e., modeled) due to the availability of multiple meta-trials. In this section, we consider several models for Σ_{tk}.

Write $S_{tk} = (S_{tk,j\ell})$ and $\Sigma_{tk} = (\Sigma_{tk,j\ell})$. Since the sample variances $S_{tk,jj}$ for $j = 1, \ldots, J$ are observed, the $\Sigma_{tk,jj}$'s are identifiable. Instead of (3.7), we consider

$$\frac{(n_{tk}-1)S_{tk,jj}}{\sum_{tk,jj}} \sim \chi^2_{n_{tk}-1}, \quad (5.1)$$

for $j = 1, \ldots, J$. Then, we assume $\Sigma_{tk,j\ell} = 0$ for $1 \le j < \ell \le J$. This model, denoted by $\mathcal{M}_1$, assumes that the $\bar{y}_{\cdot tk1}, \ldots \bar{y}_{\cdot tkJ}$ are independent. The simplest and perhaps most identifiable model is

$$\sum_{tk} = \sum. \quad (5.2)$$

This model, denoted by $\mathcal{M}_2$, assumes that all the within-study covariance matrices are the same across treatment arms and trials.

The covariance matrix Σ_{tk} can be decomposed to standard deviations and correlations as

$$\sum_{tk} = \mathrm{diag}(\sigma_{tk,11}, \ldots, \sigma_{tk,JJ}) \rho_{tk} \mathrm{diag}(\sigma_{tk,11}, \ldots, \sigma_{tk,JJ}), \quad (5.3)$$

where $\sum_{tk,jj} = \sigma^2_{tk,jj}$ for $j = 1, \ldots, J$ and ρ_{tk} is a $J \times J$ correlation matrix with diagonal elements $\rho_{tk,jj} = 1$ for $j = 1, \ldots, J$. The third model for Σ_{tk}, denoted by $\mathcal{M}_3$, assumes that

$$\rho_{tk} = \rho, \quad (5.4)$$

where $\rho = (\rho_{jj'})$ is a correlation matrix with $\rho_{jj} = 1$ for $j = 1, \ldots, J$. Model $\mathcal{M}_3$ relaxes the assumption of equal variances under model $\mathcal{M}_2$. However, both models $\mathcal{M}_2$ and $\mathcal{M}_3$ assume that the correlations among the aggregate responses, $\bar{y}_{\cdot tk1}, \ldots \bar{y}_{\cdot tkJ}$, are the same across treatment arms and trials.

Finally, we consider a hierarchical model for Σ_{tk} denoted by $\mathcal{M}_4$. This model assumes that

$$\sum_{tk}^{-1} \sim \mathrm{Wishart}_J \left(\upsilon, (\upsilon - J - 1) \sum\right). \quad (5.5)$$

In (5.5), Σ_{tk} has prior expectation $E[\Sigma_{tk}|\Sigma] = (\nu - J - 1)^{-1} (\upsilon - J - 1)\Sigma = \Sigma$ when $\upsilon > J + 1$. The hierarchical model is attractive as it allows for "borrowing of strength" across treatment arms and trials through the common second-level covariance matrix Σ, and it also accounts for the heterogeneity of the within-study covariance matrices between treatment arms as well as among different trials at the same time. We see from (5.5) that the amount of borrowing across treatment arms and trials is controlled by ν. The larger the value of ν, the more the within-study covariance matrices borrow strength from different treatment arms and trials. We consider a fixed ν in this paper and the optimal value of ν is determined by a Bayesian model assessment criterion such as the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002). One limitation of this hierarchical model for Σ_{tk} is that a large number (i.e., TK) of treatment arm and trial combinations is required in order to accurately estimate the common second-level covariance matrix Σ.

6 Bayesian Inference

6.1 Priors and Posteriors

We assume that $\boldsymbol{\beta}$, Ω_j, and Σ_{tk} are independent *a priori*. We further assume $\boldsymbol{\beta} \sim N_p(0, c_{01} I_p)$, where $p = \sum_{j=1}^{J} p_j$, and $\Omega_j^{-1} \sim \mathrm{Wishart}_{q_j} (d_{0j}, \Omega_{0j})$ with d_{0j} degrees of freedom and a $q_j \times q_j$ scale matrix Ω_{0j}, i.e.,

$$\pi\left(\Omega_j^{-1} | d_{0j}, \Omega_{0j}\right) \propto \left|\Omega_j^{-1}\right|^{\frac{d_{0j} - q_j - 1}{2}} \exp\left[-\frac{1}{2} tr\left(\Omega_{0j}^{-1} \Omega_j^{-1}\right)\right] \quad (6.1)$$

for $j = 1, \ldots, J$. As in Section 5, for Σ_{tk}, we assume

$\mathscr{M}_1$. $\sum_{tk}=\text{diag}\left(\sigma^2_{tk,11},\ldots,\sigma^2_{tk,JJ}\right)$, where $\sigma^2_{tk,jj}\sim$ Inverse Gamma (a_0,b_0) with density given by $\pi\left(\sigma^2_{tk,jj}\right)\propto\left(\sigma^2_{tk,jj}\right)^{-(a_0+1)}\exp\left(-b_0/\sigma^2_{tk,jj}\right)$, $a_0 > 0$, and $b_0 > 0$;

$\mathscr{M}_2$. $\Sigma_{tk} = \Sigma$, where $\Sigma \sim \text{Wishart}_J(\upsilon_0, \Sigma_0)$ with υ_0 degrees of freedom and $J \times J$ scale matrix Σ_0;

$\mathscr{M}_3$. $\Sigma_{tk} = \text{diag}(\sigma_{tk,11}, \ldots, \sigma_{tk,JJ})\,\rho\,\text{diag}(\sigma_{tk,11}, \ldots, \sigma_{tk,JJ})$, where $\pi(\rho)\propto 1$ subject to that ρ is a positive definite correlation matrix and $\sigma_{tk,jj}\sim$ Inverse Gamma(a_0, b_0); and

$\mathscr{M}_4$. $\sum_{tk}^{-1}\sim\text{Wishart}_J\left(\upsilon,(\upsilon-J-1)\sum\right)$ and $\Sigma\sim\text{Wishart}_J(\upsilon_0,\Sigma_0)$.

We note that c_{01}, d_{01}, …, d_{0J}, Ω_{01}, …, Ω_{0J}, a_0, b_0, υ_0, and Σ_0 are prespecified hyperparameters. In this paper, we used $c_{01} = 100{,}000$, $d_{0j} = q_j + 0.1$ and $\Omega_{0J} = 10I_{q_j}$ for $j = 1, \ldots, J$, $a_0 = 0.1$, $b_0 = 0.1$, $\upsilon_0 = J + 0.1$, and $\Sigma_0 = 10I_J$, where I_{q_j} is the $q_j \times q_j$ identity matrix for $j = 1, \ldots, J$ and I_J denotes the $J \times J$ identity matrix. We also note that in our computation development in Section 6.2 as well as Appendix B of the supplementary document, we use the sampling algorithm based on partial correlations in Section 4 to sample ρ under model $\mathscr{M}_3$ to ensure that ρ is a positive definite correlation matrix.

To obtain the posterior distributions, we consider the hierarchical model ($\mathscr{M}_4$). Let $X_{tk}=\text{diag}(\boldsymbol{x}'_{tk1},\ldots,\boldsymbol{x}'_{tkJ})$, $Z_{tk}=\text{diag}(\boldsymbol{z}'_{tk},\ldots,\boldsymbol{z}'_{tk})$, $\gamma_k=(\boldsymbol{\gamma}'_{k1},\ldots,\boldsymbol{\gamma}'_{kJ})'$, $\boldsymbol{\gamma}=(\boldsymbol{\gamma}'_1,\ldots,\boldsymbol{\gamma}'_K)'$, $R = (R_{11}, \ldots, R_{KT})$, and $\Omega = (\Omega_1, \ldots, \Omega_J)$. We further let

$$D_{obs}=\{(\overline{\boldsymbol{y}}_{.tk},n_{tk},V_{tk},\boldsymbol{X}_{tk},\boldsymbol{Z}_{tk}),t=1,\ldots,T,k=1,\ldots,K\}$$

denote the observed aggregate meta-data, where the V_{tk}'s are defined in (3.10). Then, the posterior distribution under model $\mathscr{M}_4$ is given by

$$\begin{aligned}\pi(\boldsymbol{\beta},\textstyle\sum*,\sum,\Omega,R,\boldsymbol{\gamma}|D_{obs}) \propto{}& \prod_{t=1}^{T}\prod_{k=1}^{K}|\textstyle\sum_{tk}|^{-\frac{1}{2}}\exp\left[-\frac{n_{tk}}{2}\left(\overline{\boldsymbol{y}}_{.tk}-X_{tk}\boldsymbol{\beta}-Z_{tk}\boldsymbol{\gamma}_k\right)'\sum_{tk}^{-1}\left(\overline{\boldsymbol{y}}_{.tk}-X_{tk}\boldsymbol{\beta}-Z_{tk}\boldsymbol{\gamma}_k\right)\right]\\ &\times\prod_{k=1}^{K}\prod_{j=1}^{J}|\Omega_j|^{-\frac{1}{2}}\exp\left(-\tfrac{1}{2}\boldsymbol{\gamma}'_{kj}\Omega_j^{-1}\boldsymbol{\gamma}_{kj}\right)\\ &\times\prod_{t=1}^{T}\prod_{k=1}^{K}|\textstyle\sum_{tk}|^{-\frac{n_{tk}-1}{2}}\left|V_{tk}^{\frac{1}{2}}R_{tk}V_{tk}^{\frac{1}{2}}\right|^{\frac{n_{tk}-J-2}{2}}\exp\left[-\tfrac{1}{2}tr\left((n_{tk}-1)\textstyle\sum_{tk}^{-1}V_{tk}^{\frac{1}{2}}R_{tk}V_{tk}^{\frac{1}{2}}\right)\right]\\ &\times\prod_{t=1}^{T}\prod_{k=1}^{K}|\textstyle\sum|^{\frac{\upsilon}{2}}|\textstyle\sum_{tk}|^{-\frac{\upsilon+J+1}{2}}\exp\left[-\tfrac{1}{2}tr\left((\upsilon-J-1)\textstyle\sum\sum_{tk}^{-1}\right)\right]\times\exp\left(-\frac{\boldsymbol{\beta}'\boldsymbol{\beta}}{2c_{01}}\right)\\ &\times|\textstyle\sum|^{\frac{\upsilon_0-J-1}{2}}\exp\left[-\tfrac{1}{2}tr\left(\textstyle\sum_0^{-1}\sum\right)\right]\times\prod_{j=1}^{J}|\Omega_j|^{-\frac{d_{0j}-q_j-1}{2}}\exp\left[-\tfrac{1}{2}\text{tr}\left(\Omega_{0j}^{-1}\Omega_j^{-1}\right)\right].\end{aligned}\tag{6.2}$$

The posterior distributions under the other three models can be derived in a similar fashion, and therefore the expressions of these posteriors are omitted here for brevity.

6.2 Computational Development

We present the development of the MCMC sampling algorithm only for model $\mathscr{M}_4$ as the other models have similar computational developments. Although the analytical evaluation

of the posterior distribution of ($\boldsymbol{\beta}$, $\boldsymbol{\Sigma}^*$, Σ, Ω, R, $\boldsymbol{\gamma}$) given in (6.2) is not possible, the proposed model allows us to develop an efficient MCMC sampling algorithm to sample from (6.2). The MCMC sampling algorithm requires sampling from the following conditional distributions in turn: (i) [$\boldsymbol{\beta}$, $\boldsymbol{\gamma}|\boldsymbol{\Sigma}^*$, Σ, Ω, R, D_{obs}]; (ii) [$\Omega|\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\Sigma}^*$, Σ, R, D_{obs}]; (iii) [$R|\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\Sigma}^*$, Σ, Ω, R, D_{obs}]; and (iv) [$\boldsymbol{\Sigma}^*$, $\Sigma|\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, Ω, R, D_{obs}]. For (i), we apply the collapsed Gibbs technique of Liu (1994) via the following identity

$$\left[\boldsymbol{\beta},\boldsymbol{\gamma}\middle|\sum^*,\sum,\Omega,R,D_{obs}\right]=\left[\boldsymbol{\gamma}\middle|\boldsymbol{\beta},\sum^*,\sum,\Omega,R,D_{obs}\right]\ \left[\boldsymbol{\beta}\middle|\sum^*,\sum,\Omega,R,D_{obs}\right]. \quad (6.3)$$

That is, we sample $\boldsymbol{\beta}$ after collapsing out $\boldsymbol{\gamma}$. For (iv), we again apply the collapsed Gibbs technique of Liu (1994) using the identity

$$\left[\sum^*,\sum\middle|\boldsymbol{\beta},\boldsymbol{\gamma},\Omega,R,D_{obs}\right]=\left[\sum^*\middle|\boldsymbol{\beta},\boldsymbol{\gamma},\sum,\Omega,R,D_{obs}\right]\ \left[\sum\middle|\boldsymbol{\beta},\boldsymbol{\gamma},\Omega,R,D_{obs}\right]. \quad (6.4)$$

That is, we sample Σ after collapsing out Σ^*. The technical details regarding these full conditional distributions are given in Appendix B of the supplementary document.

6.3 Bayesian Model Comparison

To carry out Bayesian model comparison, we use the Deviance Information Criteria (DIC) developed by Spiegelhalter et al. (2002). Due to the nature of the random effects, we first derive the observed-data likelihood for the model given in (3.5) and (3.6). After integrating out the random effects ($\boldsymbol{\gamma}_k$) from the complete-data likelihood function in (3.12), we have

$$\begin{aligned} L_o(\boldsymbol{\beta},\textstyle\sum^*,\Omega|D_{obs})=\prod_{k=1}^{K}\prod_{t=1}^{T}\Bigg(&\frac{\left|\sum_{tk}+n_{tk}\boldsymbol{Z}_{tk}\Omega\boldsymbol{Z}'_{tk}\right|^{-\frac{1}{2}}}{(2\pi)^{\frac{J}{2}}|n_{tk}|^{-\frac{1}{2}}}\\ &\times\exp\left\{-\tfrac{n_{tk}}{2}\ (\overline{\boldsymbol{y}}_{\cdot tk}-\boldsymbol{X}_{tk}\boldsymbol{\beta})'\left(\textstyle\sum_{tk}+n_{tk}\boldsymbol{Z}_{tk}\Omega\boldsymbol{Z}'_{tk}\right)^{-1}(\overline{\boldsymbol{y}}_{\cdot tk}-\boldsymbol{X}_{tk}\boldsymbol{\beta})\right\}\\ \times\ (n_{tk}-1)^{\frac{J(J+1)}{2}}&\left\{2^{\frac{(n_{tk}-1)J}{2}}\pi^{\frac{J(J-1)}{4}}\prod_{j=1}^{J}\left(\tfrac{n_{tk}-j}{2}\right)\right\}^{-1}|\textstyle\sum_{tk}|^{-\frac{n_{tk}-1}{2}}|(n_{tk}-1)S_{tk}|^{-\frac{n_{tk}-J-2}{2}}\\ \times\exp&\left[-\tfrac{1}{2}tr\left\{(n_{tk}-1)\textstyle\sum_{tk}^{-1}S_{tk}\right\}\right]\Bigg). \end{aligned} \quad (6.5)$$

Let $\boldsymbol{\theta}=(\boldsymbol{\beta},\boldsymbol{\Sigma}^*,\Omega)$. We define the deviance function as

$$\mathrm{Dev}(\boldsymbol{\theta})=-2\log L_{o\boldsymbol{y}}\left(\boldsymbol{\beta},\sum^*,\Omega|D_{obs}\right), \quad (6.6)$$

where

$$L_{o\boldsymbol{y}}(\boldsymbol{\beta}, \sum^{*}, \Omega | D_{obs}) = \prod_{k=1}^{K} \prod_{t=1}^{T} \left(\frac{\left| \sum_{tk} + n_{tk} \boldsymbol{Z}_{tk} \Omega \boldsymbol{Z}'_{tk} \right|^{-\frac{1}{2}}}{(2\pi)^{\frac{J}{2}} |n_{tk}|^{-\frac{1}{2}}} \right.$$
$$\left. \times \exp \left\{ -\frac{n_{tk}}{2} (\overline{\boldsymbol{y}}_{\cdot tk} - \boldsymbol{X}_{tk}\boldsymbol{\beta})' \left(\sum_{tk} + n_{tk} \boldsymbol{Z}_{tk} \Omega \boldsymbol{Z}'_{tk} \right)^{-1} (\overline{\boldsymbol{y}}_{\cdot tk} - \boldsymbol{X}_{tk}\boldsymbol{\beta}) \right\} \right). \quad (6.7)$$

According to Spiegelhalter et al. (2002), DIC is given by

$$\mathrm{DIC} = \mathrm{Dev}(\bar{\boldsymbol{\theta}}) + 2p_D, \quad (6.8)$$

where $p_D = \overline{\mathrm{Dev}(\boldsymbol{\theta})} - \mathrm{Dev}(\bar{\boldsymbol{\theta}})$, $\overline{\mathrm{Dev}(\boldsymbol{\theta})} = E[\mathrm{Dev}(\boldsymbol{\theta}) | D_{obs}]$ (the posterior mean of Dev($\boldsymbol{\theta}$)), and $\bar{\boldsymbol{\theta}} = E[\boldsymbol{\theta} | D_{obs}]$ (the posterior mean of $\boldsymbol{\theta}$). In DIC (6.8), the first term measures the goodness-of-fit, and p_D is the effective number of model parameters. The DIC in (6.8) is a Bayesian measure of fit or adequacy with $2p_D$ being the respective dimension penalty term. The smaller the DIC value, the better the model fits the data.

We note here that when the with-study sample covariance matrix S_{tk} is completely observed, model (3.7) would not be needed as long as n_{tk} is large (Yao et al., 2011). Model (3.7) is primarily used for deriving the joint distribution of the missing off-diagonal elements in S_{tk}. The DIC measure in (6.6) allows us to assess the impact of the model for Σ_{tk} on the goodness-of-fit of the meta-regression model based on (6.7).

7 Analysis of the Cholesterol Data

For the multivariate aggregate meta-data discussed in Section 2, we have $K = 26$ trials, $J = 3$, and $T = 2$ (two treatment arms, i.e., “Statin” or “Statin + EZE”). and patients in each trial were either all on statin or all not on statin prior to the trial. The sample size of the t^{th} treatment group for the k^{th} trial is n_{tk}. The values of the n_{tk}'s are shown in Figures 1 and 2 as well as in Tables A1 and A2 in Appendix A of the supplementary document. Let $\bar{\boldsymbol{y}}_{\cdot tk} = (\bar{y}_{\cdot tk1}, \bar{y}_{\cdot tk2}, \bar{y}_{\cdot tk3})'$ be the 3-dimensional mean vector of responses for the t^{th} treatment group in the k^{th} trial, where $\bar{y}_{\cdot tk1}$, $\bar{y}_{\cdot tk3}$, and $\bar{y}_{\cdot tk3}$ denote the mean percent changes in LDL-C, HDL-C, and TG from their baseline values, respectively. Also let $\mathrm{trt}_{tk} = 1$ if the i^{th} treatment group received “Statin + EZE” and 0 if “Statin” alone, and $\mathrm{onstatin}_{tk} = 1$ if the t^{th} treatment group is on statin and 0 if not on statin prior to the trial. We further let $\boldsymbol{x}_{tkj} = ((1 - \mathrm{onstatin}_{tk}), \mathrm{trt}_{tk} \times (1 - \mathrm{onstatin}_{tk}), \mathrm{onstatin}_{tk}, \mathrm{trt}_{tk} \times \mathrm{onstatin}_{tk}, (\mathrm{bl\_ldlc})_{tkj}, (\mathrm{bl\_hdlc})_{tkj}, (\mathrm{bl\_tg})_{tkj}, \mathrm{age}_{tkj}, \mathrm{white}_{tkj}, \mathrm{male}_{tkj}, \mathrm{DM}_{tkj}, \mathrm{Dur}_{tkj})'$ be the vector of covariates and $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{j,12})'$ is the vector of corresponding regression coefficients for the j^{th} response for $j = 1, 2, 3$. Then, the multivariate meta-regression model, which is a special case of the model defined in (3.5), can be written as

$$\overline{y}_{\cdot tkj} = \boldsymbol{x}'_{tkj}\boldsymbol{\beta}_j + [\gamma_{kj0} + \gamma_{kj1}\mathrm{trt}_{tk}](1 - \mathrm{onstatin}_{tk}) + [\gamma_{kj2} + \gamma_{kj3}\mathrm{trt}_{tk}]\mathrm{onstatin}_{tk} + \overline{\in}_{\cdot tkj}. \quad (7.1)$$

We write $\gamma_{kj} = (\gamma_{kj0}, \gamma_{kj1}, \gamma_{kj2}, \gamma_{kj3},)'$, which denotes the vector of random effects for the j^{th} response, and assume that Ω_j takes the form

$$\Omega_j = \begin{pmatrix} \Omega_{j00} & \Omega_{j01} & 0 & 0 \\ \Omega_{j01} & \Omega_{j11} & 0 & 0 \\ 0 & 0 & \Omega_{j22} & \Omega_{j23} \\ 0 & 0 & \Omega_{j23} & \Omega_{j33} \end{pmatrix} = \begin{pmatrix} \Omega_j^1 & 0 \\ 0 & \Omega_j^2 \end{pmatrix}. \quad (7.2)$$

For notational simplicity, we apply (6.1) only for Ω_j^1 or Ω_j^2 with $q_j = 2$ for $j = 1, \ldots, J$. In (7.2), Ω_{j00} and Ω_{j11} capture the variability of γ_{kj0} and γ_{kj1}, and Ω_{j01} captures the correlation between γ_{kj0} and γ_{kj1} among the trials in which patients were not on statin (the first line therapy); and similarly, Ω_{j22} and Ω_{j33} capture the variability of γ_{kj2} and γ_{kj3}, and Ω_{j23} captures the correlation between γ_{kj2} and γ_{kj3} among the trials in which patients were on statin (the second line therapy). The random effects γ_{kj}, which are independent of $\bar{\in}_{\cdot tkj}$, capture the heterogeneity across the K trials for the j^{th} response.

We fit the four models discussed in Sections 5 and 6.1 to the cholesterol data. We computed the DICs defined in (6.8). The values of $\text{Dev}(\bar{\boldsymbol{\theta}})$, p_D, and DIC are reported in Table 1. The results shown in Table 1 are quite interesting and intuitively appealing. First, the independence model (i.e., $\mathcal{M}_1$) has the largest DIC value (785.52), which implies that the multivariate aggregate outcome variables are indeed dependent. Second, the DIC value under the equal within-study covariance matrix model (i.e., $\mathcal{M}_2$) is smaller than the DIC value under the independence model, but much larger than those under models $\mathcal{M}_3$ and $\mathcal{M}_4$, which indicates that there is substantial heterogeneity in the within-study covariance matrices across treatment arms and trials. Third, the DIC value under model $\mathcal{M}_3$ is similar to those under $\mathcal{M}_4$ when $\upsilon \geq 10$. This result indicates that these two models fit the data equally well. However, model $\mathcal{M}_3$ is more parsimonious than model $\mathcal{M}_4$. Fourth, under model $\mathcal{M}_4$, the DIC is roughly a "convex" function of υ and the "best" DIC value is attained at $\upsilon = 10$. These results suggest that there is indeed a considerable amount of heterogeneity in the within-study covariance matrices between treatment arms and across trials.

The posterior estimates, including the posterior means, posterior standard deviations (SDs), and 95% highest posterior density (HPD) intervals of the parameters under models $\mathcal{M}_1$ to $\mathcal{M}_4$ with $\upsilon = 10$ are reported in Tables 2 to 4. We see from Tables 2 and 3 that the posterior estimates for (β_{12}, β_{14}), (β_{22}, β_{24}) and (β_{32}, β_{34}) were similar under all four models and all four models indicate that patients on "statin + EZE" had substantially higher percent changes from baseline in LDL-C, HDL-C, and TG than those on statin alone in both first and second line therapy studies (i.e., the 95% HPD intervals do not contain 0). The respective 95% HPD intervals under these four models were (-14.43, -10.02), (-14.18, -9.92), (-13.97, -9.94), and (-13.99, -9.87) for β_{12} in the first line therapy and (-22.98, -17.07), (-23.06, -17.81), (-22.70, -17.63), (-23.03, -17.81) for β_{14} in the second line therapy for the percent change from baseline in LDL-C; and (-6.91, -2.80), (-6.88, -2.79), (-6.63, -2.81), and (-6.63, -2.74) for β_{32} in the first line therapy and (-10.60, -7.19), (-10.45, -6.92), (-10.40, -7.33), (-10.46, -7.40) for β_{34} in the second line therapy for the percent change from baseline in TG. We also see substantial improvement in HDL-C from baseline for patients on "statin + EZE" over those on statin alone in both the first and second line therapy studies. The corresponding 95% HPD intervals were (1.17, 3.04), (1.17, 3.12), (1.13, 3.03),

and (0.99, 2.98) for β_{22} in the first line therapy and (0.74, 2.00), (0.64, 1.96), (0.73, 1.97), and (0.71, 1.98) for β_{24} in the second line therapy, respectively, under models $\mathcal{M}_1 - \mathcal{M}_4$.

Among these eight covariates, the trial duration regression coefficient had an HPD interval not containing zero only for the outcome variable LDL-C, and there was a substantial improvement in HDL-C from baseline for white (vs. black or hispanic) under all four models. The 95% HPD intervals for $\beta_{1,12}$ were (0.06, 2.00), (0.14, 2.22), (0.05, 1.92), and (0.09, 2.07); and the HPD intervals for β_{29} were (3.02, 8.12), (2.91, 8.44), (3.00, 8.19), and (3.17, 8.27), respectively, under $\mathcal{M}_1 - \mathcal{M}_4$. The other important covariates, which were only in TG, were gender (male versus female) under $\mathcal{M}_1 - \mathcal{M}_4$. The 95% HPD intervals for $\beta_{3,10}$ were (0.51, 26.57), (1.48, 27.85), and (2.77, 29.66), respectively, under $\mathcal{M}_2 - \mathcal{M}_4$. We also see from Table 4 that there was an important correlation ρ_{13} between LDL-C and TG under $\mathcal{M}_2 - \mathcal{M}_4$ and the corresponding 95% HPD intervals were (0.49, 0.54), (0.44, 0.98), and (0.46, 0.86), respectively. The 95% HPD for ρ_{13} under model $\mathcal{M}_3$ is wider than the corresponding intervals under models $\mathcal{M}_2$ and $\mathcal{M}_4$. Overall, the posterior estimates under models $\mathcal{M}_2 - \mathcal{M}_4$ were more similar but less similar to those of model $\mathcal{M}_1$. These results further confirm that there is a need to model these three primary outcome variables jointly.

Finally, we carried out a sensitivity analysis on the specification of hyperparameters υ_0 and Σ_0 for Σ under model $\mathcal{M}_2$. Specifically, we specified $(\upsilon_0, \Sigma_0) = (3.01, 100I_3)$ and $(\upsilon_0, \Sigma_0) = (3.0001, 10000I_3)$, which yield much more noninformative priors for Σ than the prior with $(\upsilon_0, \Sigma_0) = (3.1, 10I_3)$ specified in Section 6.1. The posterior estimates under these two priors are given in Table C1 of the supplementary document. Comparing the posterior estimates in Table C1 to those in Tables 2 and 4, we see that these posterior estimates were very similar and in particular, the posterior estimates of the correlation parameters were almost identical. These results indicate that model $\mathcal{M}_2$ is identifiable although this model did not fit the data as well as model $\mathcal{M}_3$. In Appendix B of the supplementary document, Table C2 shows the posterior estimates under model $\mathcal{M}_4$ for $\nu = 15$ and $\nu = 20$. Comparing Table C2 to Tables 3 and 4, we see that the posterior estimates for those regression coefficients, whose 95% HPD intervals do not contain 0, were quite similar while the posterior estimates of ρ_{12} were more similar than those of ρ_{13}. Thus, the posterior estimates under model $\mathcal{M}_4$ were not as robust to the specification of ν as those under model $\mathcal{M}_2$ to the specification of (υ_0, Σ_0). For this reason, we used the DIC measure to determine the optimal value of ν as discussed in Section 5.

In all of the posterior computations, we first generated 100,000 MCMC iterations with a burn-in of 20,000 iterations, and we then used 20,000 iterations obtained from every 5th iteration for computing all the posterior estimates as well as the DICs. The computer programs were written in FORTRAN 95 using IMSL subroutines with double-precision accuracy. The run-times for models $\mathcal{M}_1 - \mathcal{M}_4$ were about 10 minutes, 1 hour and 20 minutes, 2 hours and 50 minutes, and 1 hours and 30 minutes, respectively, on a Dell PC with an Intel i5 processor, 2.40 GHz CPU, and 6 GB of memory. The convergence of the MCMC sampling algorithm for all the parameters was checked following the recommendations of Cowles and Carlin (1996). All trace plots and autocorrelation plots showed good convergence and mixing of the MCMC sampling algorithm.

8 Simulation Studies

8.1 Simulation Study I

We carry out a simulation study using the partial correlation algorithm developed in Section 4. Suppose that the true 3 × 3 covariance matrix Σ is known. The diagonal elements, which are the true variances, are set to be 10 and 100. The pairwise correlations take values between 0.1 and 0.9. Thus, the off diagonal elements of Σ take values between 10 and 90. The number of patients, n, is set to be 50, 100, 500, or 1000. To initiate the simulation, we first generate the sample covariance matrix S from the Wishart distribution given in (3.9). The observed sample correlations, denoted by $r_{jj'}$ for $1 \le j < j' \le 3$, are calculated from S. Suppose for each sampled $S = V^{1/2}RV^{1/2}$, only the diagonal elements are available. We employ the partial correlation sampling algorithm at this step to draw samples from the conditional density of R given V and Σ as in (3.11). The off-diagonal elements of R that are sampled under this scenario are the so-called conditional means. We denote them by $\tilde{r_{jj'}}$ for $1 \le j > j' \le 3$.

For each simulated dataset, we generated the first 5000 iterations as burn-in then examined the convergence and performance of the partial correlation algorithm using 20,000 iterations. For illustrative purposes, we present two simulation studies here by their graphical displays. For the first simulation study shown in Figure 3, the true correlations are 0.2, 0.5 and 0.8, respectively. For the second study shown in Figure 2, all three correlations take the value of 0.9. Trace plots for both simulation studies show excellent convergence of the proposed simulation algorithm. We can also see that the 20,000 iterations make up a representative sample of the defined population with a symmetric density and a mode at the true value. For each simulation study, we generated 10,000 datasets for a given sample size and covariance matrix. Let $r_{jj'}^{(b)}$ be the observed sample correlation (OSC) and also let $\tilde{r}_{jj'}^{(b)}$ be the conditional mean (CM) from the b^{th} simulated data set, where $1 \le j \le j' \le 3$ and $b = 1, \ldots, B$. The average of the observed sample correlations is calculated as $\mathrm{AOSC} = \frac{1}{B}\sum_{b=1}^{B} r_{jj'}^{(b)}$, and the average of the conditional means is calculated as $\mathrm{ACM} = \frac{1}{B}\sum_{b=1}^{B} \tilde{r}_{jj'}^{(b)}$. We compare the observed sample correlations to the conditional means to show the power and accuracy of the sampling algorithm. The root of the average squared difference (RASD) is calculated based on the difference between the observed and conditional means of each simulation, that is, $\mathrm{RASD} = \left\{ \frac{1}{B}\sum_{b=1}^{B} \left(r_{jj'}^{(b)} - \tilde{r}_{jj'}^{(b)} \right)^2 \right\}^{1/2}$. We also compute the 95% confidence interval $\left(\tilde{r}_{jj'\,Low}^{(b)}, \tilde{r}_{jj'\,Up}^{(b)} \right)$ for each $r_{jj'}^{(b)}$. The 95% coverage conditional probability (CCP) of $\left(\tilde{r}_{jj'\,Low}^{(b)}, \tilde{r}_{jj'\,Up}^{(b)} \right)$ containing $r_{jj'}^{(b)}$ is given by $\frac{1}{B}\sum_{b=1}^{B} 1\{ \tilde{r}_{jj'\,Low}^{(b)} \le r_{jj'}^{(b)} \le \tilde{r}_{jj'\,Up}^{(b)} \}$, where the indicator function $1\{A\} = 1$ when A is true and 0 otherwise.

Table 5 summarizes the results of the simulation studies with various combinations of correlations (ρ_{12}, ρ_{23}, ρ_{13}), true variances, and sample sizes n. We used n = 50, 100, 500,

and 1000. For each sample size, we considered three different values of the correlations, which are $(\rho_{12}, \rho_{23}, \rho_{13}) = (0.1, 0.1, 0.1)$ representing small correlations; $(\rho_{12}, \rho_{23}, \rho_{13}) = (0.2, 0.5, 0.8)$ representing moderate correlations; and $(\rho_{12}, \rho_{23}, \rho_{13}) = (0.9, 0.9, 0.9)$ representing large correlations, as well as true variances equal to 10 and 100. We can see from Table 5 that the averages of the conditional means were very close to their average observed counterparts for all 18 cases. Also, the realizations of both the Wishart distribution in (3.7), which is measured by the averages of the observed sample correlations, and the conditional density for the correlation matrix in (3.11), measured by the averages of the conditional means, were close to the true correlations. All of these indicate that the sampling algorithm based on partial correlations can accurately recover the missing elements in the sample correlation matrix. The RASD calculated based on the errors between the observed and sampled correlations also represents the degree of accuracy of the sampling algorithm. We can see that when fixing the correlations $(\rho_{12}, \rho_{23}, \rho_{13})$, the RASD decreases as the sample size n increases. For example, for true variances equal to 100, when $(\rho_{12}, \rho_{23}, \rho_{13}) = (0.1, 0.1, 0.1)$, RASD = (0.139, 0.145, 0.143) for $n = 50$, RASD = (0.101, 0.101, 0.099) for $n = 100$; RASD = (0.043, 0.044, 0.044) for $n = 500$; and RASD = (0.032, 0.031, 0.031) for $n = 1000$. This is true regardless of the magnitude of the correlations. Finally, we see from Table 5 that all the values of the 95% CCP's were around 95%, indicating that the 95% confidence interval for each simulated dataset contains the observed value about 95% of the time for the 10,000 simulated datasets. Finally, we note that we also considered the case in which the true variances were set to be 1, and the results are given in Table D1 in Appendix D of the supplementary document. From Table 5 and Table D1, we see that the results for the true variances equal to 1 were very similar to those for the true variances equal to 10 or 100.

8.2 Simulation Study II

To examine the performance of the proposed method, we design a simulation study with $K = 26$ trials, $J = 3$, and $T = 2$, which mimics the cholesterol data analyzed in Section 7. The sample sizes (n_{tk}), onstatin_{tk}, trt_{tk}, age in years, duration in weeks, and the true values of the diagonal elements of Σ_{tk} for $t = 1, 2$ and $k = 1, \ldots, 26$ are chosen to be the same as those in the cholesterol data in Section 2 and analyzed in Section 7. We set $\boldsymbol{x}_{tkj} = ((1 - \text{onstatin}_{tk}), \text{trt}_{tk} \times (1 - \text{onstatin}_{tk}), \text{onstatin}_{tk}, \text{trt}_{tk} \times \text{onstatin}_{tk}, \text{age}_{tkj}, \text{Dur}_{tkj})$ to be the vector of covariates and let $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{j6})'$ be the vector of corresponding regression coefficients for the j^{th} response for $j = 1, 2, 3$. The aggregate responses, $\bar{\boldsymbol{y}}_{\cdot tk} = (\bar{y}_{\cdot tk1}, \bar{y}_{\cdot tk2}, \bar{y}_{\cdot tk3})'$, are generated from (7.1) under model $\mathcal{M}_3$ with $\boldsymbol{\rho} = (\rho_{12}, \rho_{13}, \rho_{23})' = (-0.7, 0.7, 0)'$, $\boldsymbol{\beta}_1 = (-51, -12.0, -11, -20, 1.5, 2.0)'$, $\boldsymbol{\beta}_2 = (8.0, 2.0, 3.5, 1.5, -1.0, 1.0)'$, and $\boldsymbol{\beta}_3 = (-14.0, -4.5, 1.0, -9.0, -1.0, 1.0)'$. Also let $\Omega = \left(\Omega_1^1, \Omega_1^2, \Omega_2^1, \Omega_2^2, \Omega_3^1, \Omega_3^2\right)$, where Ω_j^ℓ, $\ell = 1, 2$, $j = 1, 2, 3$, are

$$\Omega_1^1 = \begin{pmatrix} 45 & -15 \\ -15 & 11 \end{pmatrix}, \Omega_1^2 = \begin{pmatrix} 19 & -10 \\ -10 & 18 \end{pmatrix}, \Omega_2^1 = \begin{pmatrix} 1.5 & -1.7 \\ -1.7 & 2.1 \end{pmatrix}, \Omega_2^2 = \begin{pmatrix} 0.8 & -0.1 \\ -0.1 & 0.2 \end{pmatrix}, \Omega_3^1 = \begin{pmatrix} 41 & -17 \\ -17 & 8 \end{pmatrix}, \Omega_3^2 = \begin{pmatrix} 1.5 & -0.8 \\ -0.8 & 1.2 \end{pmatrix}$$

defined in (7.2). Set, . In our data generation, we also consider three sets of values, namely, Ω, Ω/2, and Ω/10, for the covariance matrix of the random effects. Using each set of the above design values, we generated 400 simulated datasets. For each simulated dataset, we fit model $\mathcal{M}_1$, model $\mathcal{M}_3$ with unknown R_{tk}, and model $\mathcal{M}_3$ with known R_{tk} since the off-diagonal elements of S_{tk} were known in the simulation setting. Fitting model $\mathcal{M}_3$ with known R_{tk} allows us to assess potential information loss due to missing off-diagonal elements of S_{tk}. In all of the posterior computations in this simulation study, we generated 20,000 MCMC iterations after a burn-in of 5,000 iterations.

Table 6 shows the average of the posterior means (EST), the simulation error (SE), which is computed as the root of the sample variance of 400 posterior means, the coverage probability (CP) of 95% HPD intervals, and the root of the mean squared error (RMSE) for each regression coefficient in $\boldsymbol{\beta}_j$, j = 1, 2, 3, over the 400 simulated datasets. In addition, we also report the mean and the interquartile range (IQR) of the 400 DIC values in the same table. From Table 6, we see that (i) the SE's and RMSE's became smaller when the covariances of the random effects were smaller for all three models we fit; (ii) the SE's and RMSE's under model $\mathcal{M}_3$ with known or unknown R_{tk}'s were either comparable to or smaller than those under model $\mathcal{M}_1$; (iii) the CP's under model $\mathcal{M}_3$ with known or unknown R_{tk}'s were closer to the desired credible level 95% than those under model $\mathcal{M}_1$, especially for Ω/2 and Ω/10; and (iv) both the means and IQRs of the DIC values under model $\mathcal{M}_1$ were larger than those under $\mathcal{M}_3$ with known or unknown R_{tk}'s. Even with one zero correlation and two moderate correlations, model $\mathcal{M}_3$ with known or unknown R_{tk}'s was more favorable than model $\mathcal{M}_1$ both in terms of the model fit and the performance of the posterior estimates of the regression coefficients. Quite interestingly, model $\mathcal{M}_3$ with unknown R_{tk} performed equally well as model $\mathcal{M}_3$ with known R_{tk} in fitting the $\bar{\boldsymbol{y}}_{\cdot tk}$ model and estimating the regression coefficients $\boldsymbol{\beta}_j$. However, the EST's, SE's, and RMSE's under model $\mathcal{M}_3$ with unknown R_{tk} were -0.67, 0.11, and 0.11 for ρ_{12}, 0.64, 0.13, and 0.14 for ρ_{13}, and 0.03, 0.16, and 0.16 for ρ_{23} when Ω was used; -0.65, 0.13, 0.14 for ρ_{12}, 0.65, 0.13, and 0.14 for p_{13}, and 0.03, 0.16, and 0.16 for p_{23} when Ω/2 was used; and -0.64, 0.12, 0.13 for p_{12}, 0.62, 0.11, and 0.14 for p_{l3}, and 0.08, 0.16, and 0.18 for ρ_{23} when Ω/10 was used. These EST's, SE's, and RMSE's under model $\mathcal{M}_3$ with known R_{tk} became -0.70, 0.0002, and 0.0002 for ρ_{12}, 0.70, 0.0002, and 0.0002 for ρ_{13}, and 0.0002, 0.0004, and 0.0004 for ρ_{23} when Ω was used; -0.70, 0.0002, 0.0002 for ρ_{12}, 0.70, 0.0002, and 0.0002 for ρ_{13}, and

0.0002, 0.0004, and 0.0004 for ρ_{23} when $\Omega/2$ was used; and -0.70, 0.0002, 0.0002 for ρ_{12}, 0.70, 0.0002, and 0.0002 for ρ_{13}, and 0.0001, 0.0004, and 0.0004 for ρ_{23} when $\Omega/10$ was used. Thus, when the sample covariance matrix S_{tk} is known, the true correlations can be very accurately recovered. These results indicate that the main information loss due to the missing off-diagonal elements in S_{tk} is in estimating $\boldsymbol{\rho}$ but there is not much information loss in estimating $\boldsymbol{\beta}_j$'s. The additional results for the meta-data generated from model $\mathcal{M}_1$ are given in Appendix D of the supplementary document.

9 Discussion

In this paper, we have proposed a novel Bayesian methodology for estimating the within-study covariance matrix in multivariate meta-regression. Our approach is based on the notion that the diagonal elements of the within-study sample covariance matrix are observed and the off-diagonal elements are treated as missing data. Then, using the Wishart distribution for the within-study sample covariance matrix, we are able to write out the complete data likelihood. Prior distributions are specified for all parameters, and a novel MCMC sampling algorithm was developed to sample from the joint posterior distribution. Our real data analyses and simulation studies were very promising. In Simulation Study I, we showed that our proposed procedure recovers (estimates) the within study correlations quite well and the MCMC algorithm converges nicely with moderate sample sizes. Our model assessment procedure also worked well for identifying the true structure of the within-study covariance matrix in terms of model fit, as shown in Simulation Study II. In the real data analysis of the cholesterol data, although we do not know the ground truth, we obtained very reasonable and interpretable results, the MCMC algorithm converged quickly and the model assessment procedure identified a reasonable model.

The development we have proposed in this paper can also be carried out within a frequentist framework using the Monte Carlo (MC) EM algorithm. An EM framework was developed for meta-regression with univariate responses in Chen et al. (2012). Thus, an extension of the methods of Chen et al. (2012) to the multivariate case is possible for the proposed model. However, such a development would require an MCEM algorithm instead of an EM algorithm. In addition, the estimated covariance matrix of the MLE's of the model parameters under multivariate responses is much more difficult and challenging to compute than the one for the univariate case. Furthermore, obtaining the MLEs of the variance components in the M-step of the MCEM algorithm would be computationally intensive and would require specialized optimization algorithms. It remains to be seen whether such a frequentist approach is computationally feasible. A Bayesian approach for this multivariate setting appears more tractable. The approach proposed in this paper for conducting multivariate meta-regression of study-level data has applications in many areas besides clinical trials. Future work in this area includes developing perhaps more general classes of models for which the response may be discrete and/or longitudinal in nature as well as the extension to the cases in which some of the aggregate responses or covariates are missing.

We implemented our methodology using the FORTRAN 95 software with double precision and IMSL subroutines. The FORTRAN 95 code is available upon request. As a future

project, we will develop a user-friendly R-interface of our already developed FORTRAN 95 software. We will make this R package available to practitioners once it is completed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank the Editor, the Associate Editor, and the anonymous reviewer for their very helpful comments and suggestions, which have led to a much improved version of the paper. Dr. Kim's research was supported by the Intramural Research Program of National Institutes of Health, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development. Dr. Chen and Dr. Ibrahim's research was partially supported by NIH grants #GM 70335 and #CA 74015.

References

Barnard J, McCulloch R, Meng XL. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. Statistica Sinica. 2000; 10:1281–1331.

Berkey CS, Anderson JJ, Hoaglin DC. Multiple-outcome metaanalysis of clinical trials. Statistics in Medicine. 1996; 15:537–557. [PubMed: 8668877]

Chen MH, Ibrahim JG, Shah AK, Lin J, Yao H. Meta-analysis methods and models with applications in evaluation of cholesterol lowering drugs. Statistics in Medicine. 2012; 31:3597–3616. [PubMed: 22829358]

Chen, MH.; Shao, QM.; Ibrahim, JG. Monte Carlo Methods in Bayesian Computation. New York: Springer; 2000.

Cowles MK, Carlin BP. Markov chain Monte Carlo convergence diagnostics: A comparative review. Journal of the American Statistical Association. 1996; 91:883–904.

Daniels MJ, Pourahmadi M. Modeling covariance matrices via partial autocorrelations. Journal of Multivariate Analysis. 2001; 100:2352–2363. [PubMed: 20161018]

Hamza TH, Arends LR, van Houwelingen HC, Stijnen T. Multivariate random effects meta-analysis of diagnostic tests with multiple thresholds. BMC Medical Research Methodology. 2009; 9:73. [PubMed: 19903336]

Hedges LV, Tipton E, Johnson MC. Robust variance estimation in meta-regression with dependent effect size estimates. Research Synthesis Methods. 2010; 1:39–65. [PubMed: 26056092]

Ibrahim JC, Chen MH, Lipsitz SR, Herring AH. Missing Data Methods in Regression Models. Journal of the American Statistical Association. 2005; 100:332–346.

Jackson D, Riley R, White IR. Multivariate meta-analysis: Potential and promise. Statistics in Medicine. 2011; 30:2481–2498. [PubMed: 21268052]

Joe H. Generating random correlation matrices based on partial correlations. Journal of Multivariate Analysis. 2006; 97:2177–2189.

Lehmann, EL.; Casella, G. Theory of Point Estimation. Second. New York: Springer; 1998.

Leiter LA, Betteridge DJ, Farmer M, Guyton JR, Lin J, Shah A, Johnson-Levonas AO, Brudi P. Lipid-altering efficacy and safety profile of combination therapy with ezetimibe/statin vs. statin monotherapy in patients with and without diabetes: an analysis of pooled data from 27 clinical trials. Diabetes, Obesity and Metabolism. 2011; 13:615–628.

Liu JS. The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. Journal of the American Statistical Association. 1994; 89:958–966.

Ma Y, Mazumdar M. Multivariate meta-analysis: a robust approach based on the theory of U-statistic. Statistics in Medicine. 2011; 30:2911–2929. [PubMed: 21830230]

Nam IS, Mengersen K, Garthwaite P. Multivariate meta-analysis. Statistics in Medicine. 2003; 22:2309–2333. [PubMed: 12854095]

Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation. Journal of the Royal Statistical Society, Series A. 2009; 172:789–811.

Riley RD, Thompson JR, Abrams KR. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. Statistics in Medicine. 2008; 9:172–186.

Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B. 2002; 64:583–639.

Stewart LA, Tierney JF. To IPD or not to IPD? Evaluation & the Health Professions. 2002; 25:76. [PubMed: 11868447]

Wang Y, Daniels MJ. Bayesian modeling of the dependence in longitudinal data via partial autocorrelations and marginal variances. Journal of Multivariate Analysis. 2013; 116:130–140. [PubMed: 23645941]

Wei Y, Higgins JPT. Estimating within-study covariances in multivariate meta-analysis with multiple outcomes. Statistics in Medicine. 2013a; 32:1191–1205. [PubMed: 23208849]

Wei Y, Higgins JPT. Bayesian multivariate meta-analysis with multiple outcomes. Statistics in Medicine. 2013b; 32:2911–2934. [PubMed: 23386217]

Whitehead, A. Meta-analysis of Controlled Clinical Trials. New York: Wiley; 2002.

Yao H, Chen MH, Qiu C. Bayesian modeling and inference for meta data with applications in efficacy evaluation of an allergic rhinitis drug. Journal of Biopharmaceutical Statistics. 2011; 21:992–1005. [PubMed: 21830927]

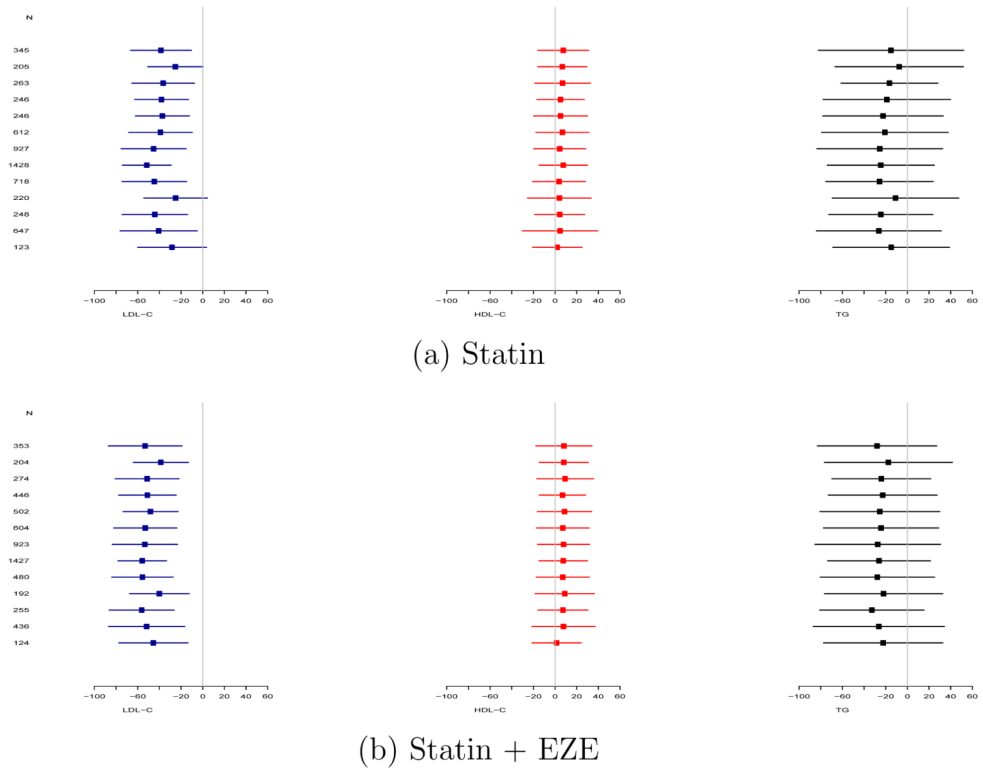


Figure 1.

Forest plots of multivariate aggregate outcome variables (LDL-C, HDL-C, TG) for patients on first line therapy.

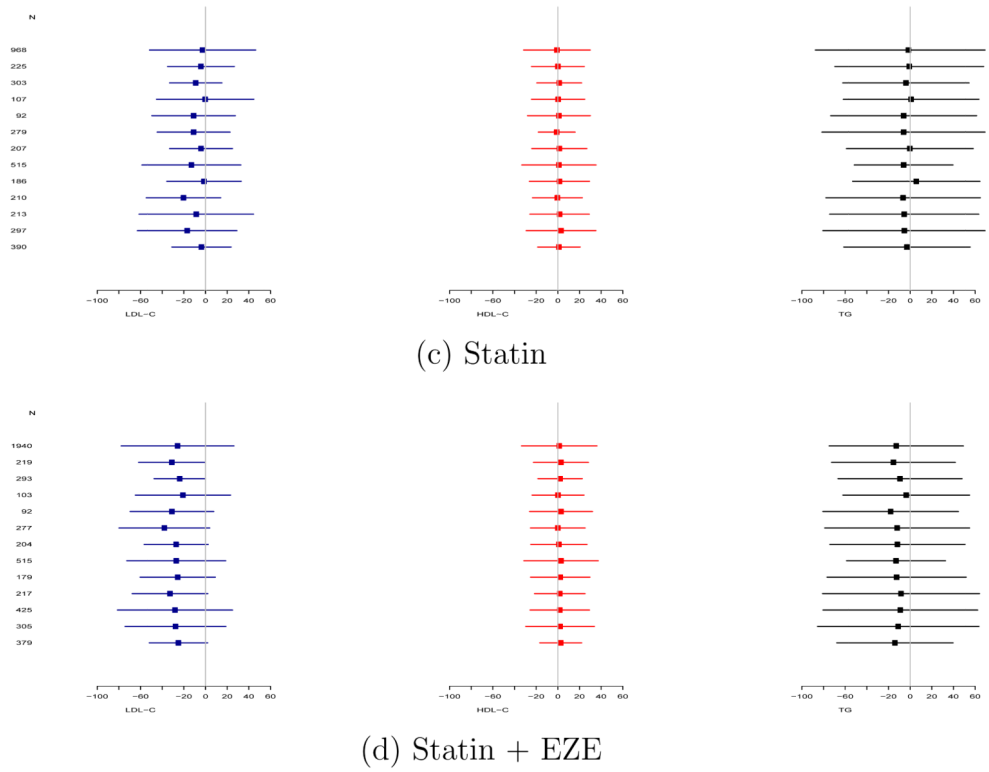
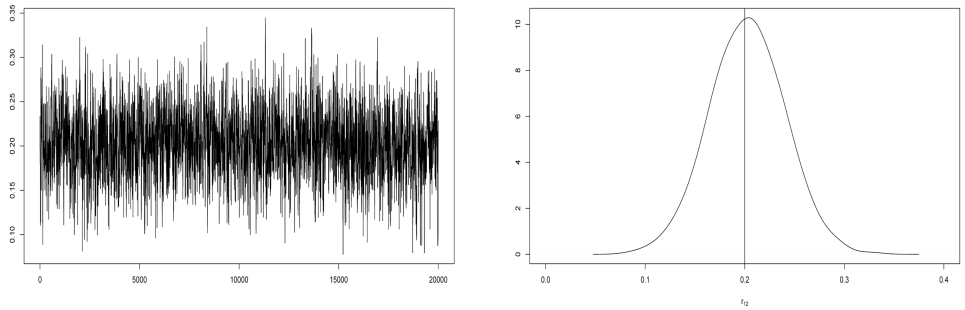
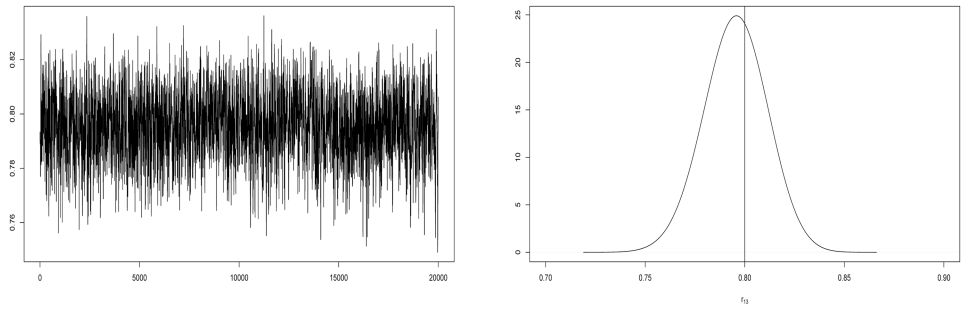


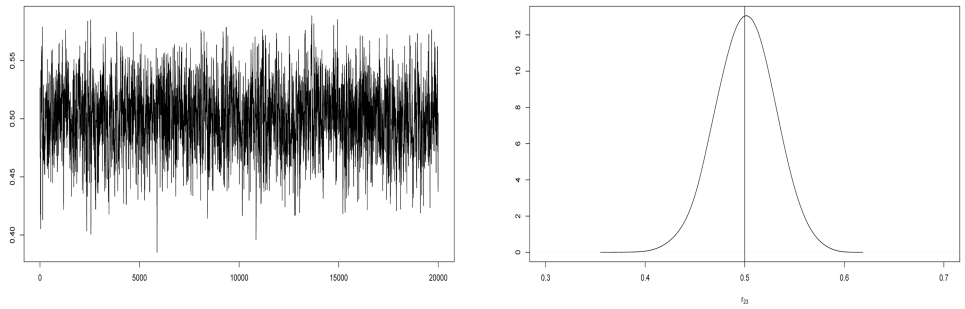
Figure 2.
Forest plots of multivariate aggregate outcome variables (LDL-C, HDL-C, TG) for patients on second line therapy.



(a1) Trace plot of r_{12} (a2) Conditional density of r_{12}



(b1) Trace plot of r_{13} (b2) Conditional density of r_{13}



(c1) Trace plot of r_{23} (c2) Conditional density of r_{23}

Figure 3.

Trace plots and density plots for simulation study with n = 500, r_{12} = 0.147, r_{12} = 0.798, r_{13} = 0.464, S_{11} = 106.4, S_{22} = 97.1, and S_{33} = 110.7

Table 1

DIC Values for various models

Model	v	Dev(θ)	p_D	DIC
$\mathcal{M}_1$		696.50	44.51	785.52
$\mathcal{M}_2$		692.41	44.37	781.15
$\mathcal{M}_3$		685.26	42.89	771.04
$\mathcal{M}_4$	5	693.60	41.59	776.77
	10	685.51	42.03	769.57
	15	687.29	42.17	771.64
	20	687.66	42.12	771.90

Table 2

Posterior Estimates of the Parameters under Models $\mathscr{M}_1$ and $\mathscr{M}_2$

			Model $\mathscr{M}_1$			Model $\mathscr{M}_2$		
			Posterior		95% HPD Interval	Posterior		95% HPD Interval
Response	**Variable**	**Parameter**	**Mean**	**SD**		**Mean**	**SD**	
LDL-C	1-onstatin	β_{11}	-64.18	51.72	(-166.75, 36.75)	-47.61	49.07	(-145.53, 49.57)
	trt × (1-ontatin)	β_{12}	-12.15	1.12	(-14.43, -10.02)	-12.03	1.08	(-14.18, -9.92)
	onstatin	β_{13}	-24.07	48.89	(-120.31, 71.56)	-6.60	46.03	(-101.26, 81.89)
	trt × ontatin	β_{14}	-20.08	1.40	(-22.98, -17.07)	-20.39	1.33	(-23.06, -17.81)
	bl_ldlc	β_{15}	0.15	0.08	(-0.02, 0.30)	0.12	0.08	(-0.04, 0.28)
	bl_hdlc	β_{16}	-0.59	0.58	(-1.69, 0.59)	-0.93	0.53	(-2.00, 0.09)
	bl_trig	β_{17}	0.004	0.08	(-0.15, 0.16)	0.03	0.09	(-0.13, 0.20)
	age	β_{18}	0.41	0.40	(-0.43, 1.16)	0.37	0.44	(-0.49, 1.24)
	white	β_{19}	-5.19	7.15	(-19.30, 8.93)	-6.53	7.35	(-21.45, 7.76)
	male	$\beta_{1,10}$	0.41	12.31	(-23.77, 24.97)	5.41	12.62	(-19.47, 29.99)
	DM	$\beta_{1,11}$	0.49	5.05	(-9.85, 9.76)	-2.25	4.52	(-11.02, 6.89)
	duration	$\beta_{1,12}$	1.04	0.50	(0.06, 2.00)	1.16	0.53	(0.14, 2.22)
	Variance of random effects	Ω_{100}	48.65	26.50	(12.62, 97.85)	41.27	23.01	(10.34, 84.66)
		Ω_{101}	-14.58	11.53	(-38.87, 3.78)	-12.57	10.13	(-32.66, 4.01)
		Ω_{111}	13.90	7.49	(3.87, 27.98)	12.12	6.76	(3.37, 24.67)
		Ω_{122}	30.67	20.92	(4.40, 68.72)	19.70	13.43	(3.65, 44.76)
		Ω_{123}	-20.96	15.27	(-51.60, 2.06)	-8.33	11.14	(-32.26, 11.31)
		Ω_{133}	25.80	13.32	(7.70, 51.78)	19.99	10.20	(6.26, 39.49)
HDL-C	1-onstatin	β_{21}	7.30	9.94	(-11.81, 27.30)	9.38	10.00	(-10.31, 28.94)
	trt × (1-ontatin)	β_{22}	2.11	0.47	(1.17, 3.04)	2.13	0.49	(1.17, 3.12)
	onstatin	β_{23}	2.89	9.63	(-15.06, 22.93)	4.74	9.73	(-14.65, 23.44)
	trt × ontatin	β_{24}	1.37	0.32	(0.74, 2.00)	1.33	0.34	(0.64, 1.96)
	bl_ldlc	β_{25}	-0.01	0.01	(-0.04, 0.02)	-0.02	0.01	(-0.04, 0.01)

			Model $\mathcal{M}_1$			Model $\mathcal{M}_2$		
			Posterior		95% HPD Interval	Posterior		95% HPD Interval
Response	**Variable**	**Parameter**	**Mean**	**SD**		**Mean**	**SD**	
	bl_hdlc	β_{26}	-0.02	0.13	(-0.27, 0.22)	-0.03	0.13	(-0.29, 0.21)
	bl_trig	β_{27}	0.02	0.02	(-0.02, 0.06)	0.02	0.02	(-0.02, 0.06)
	age	β_{28}	-0.10	0.07	(-0.25, 0.04)	-0.10	0.08	(-0.25, 0.05)
	white	β_{29}	5.55	1.31	(3.02, 8.12)	5.62	1.41	(2.91, 8.44)
	male	$\beta_{2,10}$	-1.99	2.73	(-7.60, 3.26)	-1.99	2.95	(-7.94, 3.75)
	DM	$\beta_{2,11}$	-1.02	1.03	(-3.07, 0.98)	-1.22	1.01	(-3.24, 0.73)
	duration	$\beta_{2,12}$	-0.002	0.07	(-0.14, 0.13)	-0.01	0.07	(-0.15, 0.13)
	Variance of random effects	Ω_{200}	1.41	0.98	(0.15, 3.23)	1.33	0.96	(0.06, 3.15)
		Ω_{201}	-1.59	1.07	(-3.63, -0.21)	-1.56	1.10	(-3.68, -0.14)
		Ω_{211}	1.98	1.32	(0.24, 4.46)	2.05	1.42	(0.18, 4.68)
		Ω_{222}	0.78	0.73	(0.01, 2.10)	0.86	0.80	(0.02, 2.36)
		Ω_{223}	-0.10	0.34	(-0.87, 0.49)	-0.09	0.39	(-0.93, 0.59)
		Ω_{233}	0.17	0.21	(0.01, 0.56)	0.20	0.26	(0.01, 0.66)
TG	1-onstatin	β_{31}	-10.89	26.11	(-61.23, 41.23)	-21.73	22.75	(-67.24, 22.15)
	trt × (1-ontatin)	β_{32}	-4.82	1.05	(-6.91, -2.80)	-4.74	1.04	(-6.88, -2.79)
	onstatin	β_{33}	3.46	24.76	(-45.32, 51.96)	-5.76	21.49	(-46.56, 37.66)
	trt × ontatin	β_{34}	-8.85	0.87	(-10.60, -7.19)	-8.77	0.90	(-10.45, -6.92)
	bl_ldlc	β_{35}	0.02	0.04	(-0.07, 0.10)	0.04	0.04	(-0.04, 0.11)
	bl_hdlc	β_{36}	-0.003	0.28	(-0.53, 0.56)	-0.02	0.24	(-0.48, 0.46)
	bl_trig	β_{37}	-0.08	0.05	(-0.18, 0.02)	-0.06	0.05	(-0.16, 0.03)
	age	β_{38}	-0.14	0.22	(-0.56, 0.29)	-0.05	0.20	(-0.46, 0.34)
	white	β_{39}	-0.37	5.12	(-10.86, 9.36)	-1.37	4.82	(-10.77, 8.05)
	male	$\Omega_{3,10}$	14.49	7.25	(-0.08, 28.55)	14.11	6.59	(0.51, 26.57)
	DM	$\Omega_{3,11}$	5.12	2.94	(-0.30, 11.20)	5.13	2.64	(-0.08, 10.31)
	duration	$\Omega_{3,12}$	0.12	0.25	(-0.36, 0.62)	0.25	0.23	(-0.21, 0.70)
	Variance of random effects	Ω_{300}	45.74	25.88	(9.80, 94.00)	39.57	22.48	(9.39, 81.99)

Response	Variable	Parameter	Model $\mathscr{M}_1$			Model $\mathscr{M}_2$		
			Posterior		95% HPD Interval	Posterior		95% HPD Interval
			Mean	SD		Mean	SD	
		Ω_{301}	-19.96	12.47	(-45.17, -2.91)	-17.98	11.35	(-39.09, -1.94)
		Ω_{311}	9.41	6.92	(0.02, 22.41)	8.77	6.60	(0.25, 21.04)
		Ω_{322}	3.38	5.11	(0.01, 13.38)	3.53	4.54	(0.004, 11.85)
		Ω_{323}	-2.24	4.23	(-11.16, 1.86)	-3.11	4.50	(-11.53, 1.07)
		Ω_{333}	2.55	4.62	(0.01, 11.17)	3.69	5.64	(0.01, 13.67)

Table 3

Posterior Estimates of the Parameters under Models $\mathcal{M}_3$ and $\mathcal{M}_4$

			Model $\mathcal{M}_3$			Model $\mathcal{M}_4$		
			Posterior		95% HPD Interval	Posterior		95% HPD Interval
Response	Variable	Parameter	Mean	SD		Mean	SD	
LDL-C	1-onstatin	β_{11}	-51.23	44.06	(-136.21, 36.53)	-46.11	47.19	(-137.51, 48.80)
	trt × (l-ontatin)	β_{12}	-11.87	1.02	(-13.97, -9.94)	-11.90	1.05	(-13.99, -9.87)
	onstatin	β_{13}	-10.98	41.55	(-95.41, 67.40)	-5.84	44.42	(-93.44, 81.65)
	trt × ontatin	β_{14}	-20.21	1.29	(-22.70, -17.63)	-20.45	1.31	(-23.03, -17.81)
	bl_ldlc	β_{15}	0.12	0.07	(-0.03, 0.26)	0.11	0.08	(-0.06, 0.26)
	bl_hdlc	β_{16}	-0.76	0.50	(-1.76, 0.21)	-0.86	0.51	(-1.85, 0.15)
	bl_trig	β_{17}	0.03	0.08	(-0.12, 0.18)	0.02	0.08	(-0.14, 0.18)
	age	β_{18}	0.35	0.39	(-0.46, 1.08)	0.35	0.41	(-0.48, 1.13)
	white	β_{9}	-5.83	6.73	(-18.75, 7.67)	-5.68	7.06	(-19.42, 8.33)
	male	$\beta_{1,10}$	2.28	11.40	(-21.62, 23.43)	3.93	12.33	(-18.97, 29.18)
	DM	$\beta_{1,11}$	-1.75	4.25	(-9.68, 6.91)	-1.96	4.39	(-10.82, 6.31)
	Dur	$\beta_{1,12}$	0.99	0.47	(0.05, 1.92)	1.07	0.50	(0.09, 2.07)
	Variance of random effects	Ω_{100}	44.31	23.73	(12.39, 89.99)	44.06	23.73	(12.13, 89.87)
		Ω_{101}	-14.40	10.11	(-35.75, 0.67)	-13.64	9.94	(-33.95, 2.14)
		Ω_{111}	11.33	6.01	(3.16, 22.44)	11.83	6.26	(3.39, 23.68)
		Ω_{122}	18.86	13.44	(2.84, 44.12)	19.42	13.31	(3.44, 44.69)
		Ω_{123}	-10.09	10.21	(-31.10, 7.14)	-9.06	10.85	(-32.50, 10.38)
		Ω_{133}	18.34	9.56	(5.69, 36.21)	19.44	9.55	(6.27, 38.15)
HDL-C	1-onstatin	β_{21}	8.01	10.15	(-11.36, 28.30)	9.91	10.31	(-10.01, 30.33)
	trt × (1-ontatin)	β_{22}	2.08	0.48	(1.13, 3.03)	2.01	0.50	(0.99, 2.98)
	onstatin	β_{23}	3.55	9.85	(-15.18, 23.44)	5.37	10.01	(-13.57, 25.52)
	trt × ontatin	β_{24}	1.36	0.32	(0.73, 1.97)	1.34	0.32	(0.71, 1.98)
	bl_ldlc	β_{25}	-0.01	0.01	(-0.04, 0.01)	-0.02	0.01	(-0.04, 0.01)

			Model $\mathcal{M}_3$			Model $\mathcal{M}_4$		
			Posterior		95% HPD Interval	Posterior		95% HPD Interval
Response	**Variable**	**Parameter**	**Mean**	**SD**		**Mean**	**SD**	
	bl_hdlc	β_{26}	-0.03	0.13	(-0.28, 0.22)	-0.03	0.13	(-0.29, 0.21)
	bl_trig	β_{27}	0.02	0.02	(-0.02, 0.06)	0.02	0.02	(-0.02, 0.07)
	age	β_{28}	-0.11	0.08	(-0.25, 0.04)	-0.13	0.08	(-0.28, 0.02)
	white	β_{29}	5.60	1.33	(3.00, 8.19)	5.71	1.30	(3.17, 8.27)
	male	$\beta_{2,10}$	-1.90	2.79	(-7.41, 3.60)	-1.57	2.86	(-7.15, 4.07)
	DM	$\beta_{2,11}$	-1.05	1.04	(-3.08, 1.00)	-1.03	1.02	(-3.00, 0.99)
	Dur	$\beta_{2,12}$	-0.01	0.07	(-0.14, 0.14)	-0.01	0.07	(-0.15, 0.13)
	Variance of random effects	Ω_{200}	1.45	1.00	(0.16, 3.37)	1.46	1.01	(0.14, 3.40)
		Ω_{201}	-1.65	1.10	(-3.76, -0.18)	-1.78	1.18	(-4.02, -0.19)
		Ω_{211}	2.07	1.37	(0.29, 4.70)	2.35	1.54	(0.31, 5.27)
		Ω_{222}	0.84	0.75	(0.02, 2.23)	0.77	0.68	(0.01, 2.03)
		Ω_{223}	-0.09	0.36	(-0.84, 0.55)	0.12	0.31	(-0.51, 0.78)
		Ω_{233}	0.17	0.22	(0.01, 0.56)	0.20	0.24	(0.01, 0.65)
TG	1-onstatin	β_{31}	-13.98	24.23	(-60.38, 34.82)	-15.82	24.48	(-63.59, 33.24)
	trt × (1-ontatin)	β_{32}	-4.63	0.97	(-6.63, -2.81)	-4.66	0.99	(-6.63, -2.74)
	onstatin	β_{33}	1.13	23.11	(-42.52, 48.42)	-0.89	23.29	(-46.58, 45.33)
	trt × ontatin	β_{34}	-8.90	0.78	(-10.40, -7.33)	-8.89	0.78	(-10.46, -7.40)
	bl_ldlc	β_{35}	0.02	0.04	(-0.06, 0.10)	0.02	0.04	(-0.06, 0.10)
	bl_hdlc	β_{36}	-0.02	0.26	(-0.55, 0.48)	0.01	0.26	(-0.50, 0.52)
	bl_trig	β_{37}	-0.07	0.05	(-0.17, 0.02)	-0.07	0.05	(-0.17, 0.02)
	age	β_{38}	-0.10	0.20	(-0.49, 0.30)	-0.10	0.20	(-0.51, 0.29)
	white	β_{39}	-1.41	4.82	(-10.80, 8.23)	-1.36	4.86	(-11.19, 8.06)
	male	$\beta_{3,10}$	15.18	6.70	(1.48, 27.85)	16.06	6.83	(2.77, 29.66)
	DM	$\beta_{3,11}$	4.57	2.73	(-0.69, 10.00)	5.08	2.68	(-0.18, 10.40)
	Dur	$\beta_{3,12}$	0.19	0.24	(-0.26, 0.67)	0.19	0.23	(-0.28, 0.63)
	Variance of random effects	Ω_{300}	40.66	23.19	(9.61, 84.10)	42.13	24.25	(10.28, 86.70)

Response	Variable	Parameter	Model $\mathscr{M}_3$			Model $\mathscr{M}_4$		
			Posterior		95% HPD Interval	Posterior		95% HPD Interval
			Mean	SD		Mean	SD	
		Ω_{301}	-17.16	11.06	(-38.68, -1.27)	-17.74	11.45	(-39.88, -2.27)
		Ω_{311}	7.86	6.05	(0.01, 18.64)	8.08	6.17	(0.03, 19.41)
		Ω_{322}	1.48	3.01	(0.01, 6.85)	1.45	2.76	(0.01, 6.55)
		Ω_{323}	-0.79	2.40	(-5.58, 1.51)	-0.69	2.29	(-5.48, 1.87)
		Ω_{333}	1.17	2.70	(0.01, 5.66)	1.27	2.81	(0.01, 5.93)

Table 4

Posterior Estimates of Covariances and/or Correlations under Models $\mathcal{M}_2$, $\mathcal{M}_3$, and $\mathcal{M}_4$

Model	Parameter	Mean	SD	95% HPD interval
$\mathcal{M}_2$	Σ_{11}	319.64	3.10	(313.58, 325.77)
Covariance	Σ_{22}	181.59	1.75	(178.17, 185.00)
	Σ_{33}	899.54	8.63	(882.80, 916.17)
Correlation	ρ_{12}	-0.66	0.01	(-0.69, -0.64)
	ρ_{13}	0.51	0.01	(0.49, 0.54)
	ρ_{23}	0.04	0.06	(-0.07, 0.16)
$\mathcal{M}_3$	ρ_{12}	-0.10	0.26	(-0.58, 0.41)
Correlation	ρ_{13}	0.78	0.16	(0.44, 0.98)
	ρ_{23}	-0.11	0.16	(-0.42, 0.21)
$\mathcal{M}_4$	Σ_{11}	288.31	20.61	(248.56, 328.96)
Covariance	Σ_{22}	190.07	13.51	(164.01, 216.60)
	Σ_{33}	885.86	60.73	(765.75, 1001.86)
Correlation	ρ_{12}	-0.68	0.16	(-0.90, -0.41)
	ρ_{13}	0.68	0.11	(0.46, 0.86)
	ρ_{23}	-0.21	0.16	(-0.52, 0.10)

Table 5

Summary of Simulation Study I with True Variances Equal to 10 and 100

		True Variances Equal to 10				True Variances Equal to 100			
n	**True Value**	**AOSC**	**ACM**	**RASD**	**95% CCP**	**AOSC**	**ACM**	**RASD**	**95% CCP**
50	$\rho 12 = 0.1$	0.096	0.103	0.140	95.9	0.098	0.103	0.139	95.9
	$\rho 23 = 0.1$	0.100	0.103	0.146	95.1	0.099	0.103	0.145	94.2
	$\rho 13 = 0.1$	0.098	0.099	0.144	94.9	0.098	0.099	0.143	95.0
	$\rho 12 = 0.2$	0.194	0.212	0.121	94.5	0.190	0.212	0.120	95.3
	$\rho 23 = 0.5$	0.497	0.510	0.092	95.2	0.493	0.509	0.094	95.8
	$\rho 13 = 0.8$	0.795	0.796	0.041	93.9	0.794	0.795	0.041	94.3
	$\rho 12 = 0.9$	0.899	0.903	0.021	95.6	0.898	0.903	0.022	93.6
	$\rho 23 = 0.9$	0.898	0.903	0.020	96.2	0.897	0.903	0.021	95.6
	$\rho 13 = 0.9$	0.898	0.900	0.021	95.2	0.898	0.900	0.021	95.2
100	$\rho 12 = 0.1$	0.099	0.101	0.101	95.7	0.099	0.101	0.101	95.4
	$\rho 23 = 0.1$	0.099	0.101	0.101	94.7	0.099	0.101	0.101	94.8
	$\rho 13 = 0.1$	0.095	0.100	0.099	94.9	0.098	0.100	0.099	95.3
	$\rho 12 = 0.2$	0.196	0.207	0.082	95.1	0.203	0.209	0.086	93.8
	$\rho 23 = 0.5$	0.497	0.506	0.063	95.4	0.501	0.506	0.065	94.3
	$\rho 3 = 0.8$	0.798	0.799	0.028	95.1	0.798	0.800	0.029	94.2
	$\rho 12 = 0.9$	0.899	0.902	0.014	95.1	0.899	0.902	0.015	94.6
	$\rho 23 = 0.9$	0.899	0.902	0.015	94.4	0.899	0.902	0.015	95.5
	$\rho 13 = 0.9$	0.899	0.900	0.014	94.6	0.900	0.901	0.014	95.4
500	$\rho 12 = 0.1$	0.100	0.100	0.043	94.8	0.100	0.101	0.043	94.9
	$\rho 23 = 0.1$	0.098	0.100	0.044	95.3	0.099	0.100	0.044	95.4
	$\rho 13 = 0.1$	0.100	0.100	0.044	94.6	0.100	0.100	0.044	94.9
	$\rho 12 = 0.2$	0.199	0.201	0.039	94.3	0.199	0.201	0.037	94.6
	$\rho 23 = 0.5$	0.499	0.501	0.030	94.0	0.499	0.500	0.029	94.5

		True Variances Equal to 10				True Variances Equal to 100			
n	**True Value**	**AOSC**	**ACM**	**RASD**	**95% CCP**	**AOSC**	**ACM**	**RASD**	**95% CCP**
	$\rho 13 = 0.8$	0.799	0.800	0.013	94.8	0.799	0.800	0.013	94.3
	$\rho 12 = 0.9$	0.900	0.900	0.007	94.1	0.900	0.901	0.006	93.9
	$\rho 23 = 0.9$	0.900	0.900	0.006	94.8	0.900	0.901	0.006	95.7
	$\rho 13 = 0.9$	0.900	0.900	0.006	95.6	0.900	0.900	0.006	95.4
1000	$\rho 12 = 0.1$	0.099	0.100	0.031	95.1	0.098	0.100	0.032	95.0
	$\rho 23 = 0.1$	0.100	0.100	0.030	95.7	0.100	0.100	0.031	95.6
	$\rho 13 = 0.1$	0.099	0.100	0.032	94.8	0.099	0.100	0.031	95.2
	$\rho 12 = 0.2$	0.200	0.201	0.026	95.5	0.201	0.201	0.027	93.5
	$\rho 23 = 0.5$	0.500	0.500	0.019	96.2	0.501	0.501	0.020	96.1
	$\rho 13 = 0.8$	0.800	0.800	0.009	94.7	0.800	0.800	0.009	94.8
	$\rho 12 = 0.9$	0.900	0.900	0.005	94.1	0.900	0.900	0.005	94.7
	$\rho 23 = 0.9$	0.900	0.900	0.005	93.9	0.900	0.900	0.004	95.2
	$\rho 13 = 0.9$	0.900	0.900	0.004	95.6	0.900	0.900	0.004	95.4

Table 6

Summary of the Posterior Estimates under Models $\mathcal{M}_1$, $\mathcal{M}_3$ with unknown R_{tk}, and $\mathcal{M}_3$ with known R_{tk} (True model: Model $\mathcal{M}_3$ with $\rho = (-0.7, 0.7, 0)$)

			Ω				Ω/2				Ω/10			
Model	**Parameter**	**True Value**	**EST**	**SE**	**95% CP**	**RMSE**	**EST**	**SE**	**95% CP**	**RMSE**	**EST**	**SE**	**95% CP**	**RMSE**
$\mathcal{M}_1$	$\beta 11$	-51.0	-49.25	14.25	93.25	14.33	-49.63	10.59	90.25	10.66	-50.72	4.85	93.00	4.85
	$\beta 12$	-12.0	-11.96	0.95	94.00	0.95	-11.97	0.71	92.50	0.71	-11.99	0.40	93.50	0.40
	$\beta 13$	-11.0	-8.88	14.78	92.50	14.91	-9.35	10.95	89.75	11.06	-10.60	5.03	92.25	5.04
	$\beta 14$	-20.0	-20.09	1.22	92.50	1.23	-20.07	0.92	90.00	0.93	-20.05	0.56	90.75	0.56
	$\beta 15$	1.5	1.47	0.23	92.75	0.24	1.48	0.17	89.75	0.17	1.49	0.08	93.50	0.08
	$\beta 16$	2.0	2.00	0.34	90.50	0.34	2.00	0.25	90.75	0.25	2.00	0.12	95.50	0.12
	$\beta 21$	8.0	7.82	3.56	90.25	3.56	7.86	2.92	92.25	2.92	7.91	2.17	96.75	2.17
	$\beta 22$	2.0	1.97	0.50	90.00	0.50	1.98	0.40	89.75	0.40	1.99	0.28	94.00	0.28
	$\beta 23$	3.5	3.30	3.76	90.25	3.76	3.34	3.08	92.00	3.09	3.40	2.30	96.75	2.30
	$\beta 24$	1.5	1.51	0.31	96.75	0.31	1.51	0.29	97.25	0.29	1.51	0.28	97.00	0.28
	$\beta 25$	-1.0	-1.00	0.06	89.50	0.06	-1.00	0.05	91.00	0.05	-1.00	0.04	96.25	0.04
	$\beta 26$	1.0	1.00	0.06	92.75	0.06	1.00	0.05	93.75	0.05	1.00	0.04	96.50	0.04
	$\beta 31$	-14.0	-14.16	6.28	91.75	6.28	-14.16	5.28	95.00	5.27	-14.16	4.43	97.50	4.43
	$\beta 32$	-4.5	-4.42	0.98	91.25	0.98	-4.44	0.80	92.25	0.80	-4.47	0.59	93.75	0.59
	$\beta 33$	1.0	1.01	6.21	92.00	6.20	0.98	5.29	95.50	5.29	0.94	4.58	98.50	4.58
	$\beta 34$	-9.0	-9.07	0.73	93.50	0.73	-9.07	0.69	94.50	0.69	-9.08	0.66	96.25	0.66
	$\beta 35$	-1.0	-1.00	0.10	93.25	0.10	-1.00	0.08	95.50	0.08	-1.00	0.07	98.25	0.07
	$\beta 36$	1.0	1.01	0.19	95.75	0.19	1.00	0.17	95.75	0.17	1.00	0.13	95.00	0.13
	DIC: Mean (IQR)		737.18 (28.79)				678.67 (27.33)				571.10 (28.53)			
$\mathcal{M}_3$ with unknown R_{tk}	$\beta 11$	-51.0	-49.51	13.42	95.75	13.49	-49.85	9.70	95.75	9.76	-50.54	4.55	94.25	4.57
	$\beta 12$	-12.0	-11.97	0.94	95.00	0.94	-11.98	0.70	93.75	0.70	-11.99	0.39	94.25	0.39
	$\beta 13$	-11.0	-9.14	13.98	94.50	14.08	-9.58	10.09	94.00	10.17	-10.41	4.73	94.25	4.76
	$\beta 14$	-20.0	-20.10	1.20	94.75	1.20	-20.07	0.89	95.00	0.89	-20.05	0.54	93.25	0.54
	$\beta 15$	1.5	1.47	0.22	95.00	0.22	1.48	0.16	94.25	0.16	1.49	0.07	94.25	0.07
	$\beta 16$	2.0	2.01	0.34	92.25	0.34	2.01	0.24	91.25	0.24	2.00	0.12	94.00	0.12

			Ω				Ω/2				Ω/10			
Model	**Parameter**	**True Value**	**EST**	**SE**	**95% CP**	**RMSE**	**EST**	**SE**	**95% CP**	**RMSE**	**EST**	**SE**	**95% CP**	**RMSE**
	$\beta 21$	8.0	7.87	3.60	89.50	3.60	7.90	2.95	92.25	2.95	7.94	2.19	97.00	2.19
	$\beta 22$	2.0	1.98	0.50	90.50	0.50	1.98	0.40	91.00	0.40	1.99	0.28	95.25	0.28
	$\beta 23$	3.5	3.36	3.80	89.75	3.80	3.39	3.12	91.50	3.12	3.43	2.32	96.25	2.32
	$\beta 24$	1.5	1.51	0.31	96.25	0.31	1.51	0.29	97.25	0.29	1.51	0.28	97.50	0.28
	$\beta 25$	-1.0	-1.00	0.06	90.25	0.06	-1.00	0.05	91.50	0.05	-1.00	0.04	96.50	0.04
	$\beta 26$	1.0	1.00	0.06	93.50	0.06	1.00	0.05	93.75	0.05	1.00	0.04	97.00	0.04
	$\beta 31$	-14.0	-14.07	6.32	90.75	6.32	-14.10	5.28	94.50	5.27	-14.07	4.39	97.00	4.39
	$\beta 32$	-4.5	-4.43	0.98	92.00	0.98	-4.44	0.79	92.25	0.80	-4.47	0.59	93.75	0.59
	$\beta 33$	1.0	1.10	6.25	90.75	6.24	1.04	5.29	95.75	5.28	1.02	4.54	98.00	4.54
	$\beta 34$	-9.0	-9.06	0.74	93.00	0.74	-9.06	0.69	94.25	0.70	-9.06	0.67	96.00	0.67
	$\beta 35$	-1.0	-1.00	0.10	92.75	0.10	-1.00	0.08	95.50	0.08	-1.00	0.07	97.75	0.07
	$\beta 36$	1.0	1.01	0.19	94.50	0.19	1.00	0.17	95.75	0.17	1.00	0.12	95.50	0.12
	DIC: Mean (IQR)		730.77 (27.05)				666.90 (26.45)				543.45 (27.26)			
$\mathcal{W}_3$ with known R_{tk}	$\beta 11$	-51.0	-49.48	13.43	95.25	13.50	-49.88	9.61	95.75	9.66	-50.51	4.56	94.00	4.58
	$\beta 12$	-12.0	-11.98	0.94	95.50	0.94	-11.98	0.70	94.50	0.70	-11.99	0.39	93.50	0.39
	$\beta 13$	-11.0	-9.12	13.99	94.50	14.10	-9.61	10.00	94.75	10.09	-10.38	4.76	95.25	4.79
	$\beta 14$	-20.0	-20.10	1.20	95.50	1.20	-20.08	0.89	94.75	0.89	-20.05	0.53	94.50	0.53
	$\beta 15$	1.5	1.47	0.22	94.50	0.22	1.48	0.16	94.00	0.16	1.49	0.07	95.00	0.07
	$\beta 16$	2.0	2.01	0.34	92.50	0.34	2.01	0.24	92.75	0.24	2.00	0.12	92.75	0.12
	$\beta 21$	8.0	7.88	3.59	90.25	3.59	7.91	2.95	92.00	2.95	7.96	2.20	96.50	2.20
	$\beta 22$	2.0	1.98	0.50	90.50	0.50	1.98	0.40	90.50	0.40	1.99	0.28	95.25	0.28
	$\beta 23$	3.5	3.38	3.79	89.75	3.79	3.41	3.11	91.75	3.11	3.46	2.33	97.00	2.33
	$\beta 24$	1.5	1.51	0.31	96.25	0.31	1.51	0.29	97.00	0.29	1.51	0.28	98.00	0.28
	$\beta 25$	-1.0	-1.00	0.06	89.50	0.06	-1.00	0.05	91.00	0.05	-1.00	0.04	96.25	0.04
	$\beta 26$	1.0	1.00	0.06	93.00	0.06	1.00	0.05	94.25	0.05	1.00	0.04	97.25	0.04
	$\beta 31$	-14.0	-14.10	6.31	91.00	6.30	-14.09	5.28	94.25	5.28	-14.06	4.40	97.00	4.40
	$\beta 32$	-4.5	-4.43	0.98	91.75	0.98	-4.44	0.79	92.25	0.80	-4.47	0.59	93.00	0.59
	$\beta 33$	1.0	1.07	6.24	91.25	6.23	1.04	5.31	95.25	5.30	1.02	4.56	97.50	4.56

			Ω				Ω/2				Ω/10			
Model	**Parameter**	**True Value**	**EST**	**SE**	**95% CP**	**RMSE**	**EST**	**SE**	**95% CP**	**RMSE**	**EST**	**SE**	**95% CP**	**RMSE**
	$\beta 34$	-9.0	-9.06	0.73	93.25	0.73	-9.06	0.69	94.50	0.69	-9.06	0.67	95.75	0.67
	$\beta 35$	-1.0	-1.00	0.10	93.00	0.10	-1.00	0.08	95.75	0.08	-1.00	0.07	97.75	0.07
	$\beta 36$	1.0	1.01	0.19	94.75	0.19	1.00	0.16	95.75	0.16	1.00	0.12	94.75	0.12
	DIC: Mean (IQR)		730.81 (27.81)				666.71 (26.12)				543.65 (25.39)			