Bayesian inference for network meta-regression using multivariate random effects with applications to cholesterol lowering drugs

HAO LI, MING-HUI CHEN

Department of Statistics, University of Connecticut, Storrs, CT 06269, USA

JOSEPH G. IBRAHIM*

Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599, USA

ibrahim@bios.unc.edu

SUNGDUK KIM

Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD 20892, USA

ARVIND K. SHAH, JIANXIN LIN, ANDREW M. TERSHAKOVEC

Merck Research Laboratories, Merck & Co., Inc., Rahway, NJ 07065, USA

SUMMARY

Low-density lipoprotein cholesterol (LDL-C) has been identified as a causative factor for atherosclerosis and related coronary heart disease, and as the main target for cholesterol- and lipid-lowering therapy. Statin drugs inhibit cholesterol synthesis in the liver and are typically the first line of therapy to lower elevated levels of LDL-C. On the other hand, a different drug, Ezetimibe, inhibits the absorption of cholesterol by the small intestine and provides a different mechanism of action. Many clinical trials have been carried out on safety and efficacy evaluation of cholesterol lowering drugs. To synthesize the results from different clinical trials, we examine treatment level (aggregate) network meta-data from 29 double-blind, randomized, active, or placebo-controlled statins +/- Ezetimibe clinical trials on adult treatment-naïve patients with primary hypercholesterolemia. In this article, we propose a new approach to carry out Bayesian inference for arm-based network meta-regression. Specifically, we develop a new strategy of grouping the variances of random effects, in which we first formulate possible sets of the groups of the treatments based on their clinical mechanisms of action and then use Bayesian model comparison criteria to select the best set of groups. The proposed approach is especially useful when some treatment arms are involved in only a single trial. In addition, a Markov chain Monte Carlo sampling algorithm is developed to carry out the posterior computations. In particular, the correlation matrix is generated from its full conditional distribution via partial correlations. The proposed methodology is further applied to analyze the network meta-data from 29 trials with 11 treatment arms.

Keywords: Bayesian cumulative ranking curve; Deviance information criteria; Localized metropolis algorithm; Logarithm of the Pseudo marginal likelihood; Partial correlations.

*To whom correspondence should be addressed.

1. INTRODUCTION

According to the National Center for Health Statistics, high cholesterol is a risk factor for heart disease, which is the leading cause of death for both men and women. Nearly 600 000 people die of heart disease in the United States every year—that's one in every four deaths. Every year about 715 000 Americans have a heart attack and the costs of coronary heart disease alone costs the US over \$100 billion annually which includes the cost of health care services, medications, and lost productivity (NCHStats, 2013). High cholesterol is well known to contribute to heart disease and other cardiovascular diseases. [Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults \(2001\)](#) has issued treatment guidelines identifying low-density lipoprotein cholesterol (LDL-C, “bad” cholesterol) as a causative factor for coronary heart disease and as the main target for cholesterol-lowering and lipid-lowering therapy.

Cholesterol-lowering medicines called “statins” work mainly in the liver to decrease the production of cholesterol and reduce cholesterol in the bloodstream. Many clinical trials since the introduction of statin drugs in 1987 have shown statins to drive down the rates of heart attack and stroke. Statin drugs are typically the first choice to drive down elevated levels of LDL-C. An estimated 38.7 million Americans were on a statin in 2011–2012, an increase from 24 million in 2003–2004, and from 12.5 million in 1990–2000 ([Adedinsewo and others, 2016](#)). There are several effective brands on generic labels and thus are quite inexpensive. There are several classes of cholesterol-lowering drugs, but statins have become the drug class of choice because of their demonstrated efficacy and safety. This class of drugs includes: atorvastatin (Lipitor), simvastatin (Zocor), lovastatin (Mevacor), rosuvastatin (Crestor), pravastatin (Pravachol), and fluvastatin (Lescol).

Between 2000 and 2014, the fraction of Americans with elevated blood levels of cholesterol declined from 18.3% to 11%, at least partly due to an increase in the use of cholesterol-lowering medications. The use of any lipid-lowering agent was 20% in 2004 and the data from 2012 showed that 28% of Americans over the age of 40 are taking such a medication ([Ross, 2015](#)). But a high blood level of LDL-C remains a major risk factor for heart disease and stroke in the US. In general, statins positively affect the lipid profile by decreasing LDL-C and triglycerides (TG), and increasing high-density lipoprotein cholesterol (HDL-C, ‘good’ cholesterol), although the effect on HDL-C is relatively smaller. Clinical studies have shown that statins significantly reduce the risk of heart attack and death in patients with coronary artery disease and can also reduce cardiac events in patients with high cholesterol levels. Even though statins are the first-line treatment in most patients, lipid goals are frequently not achieved because of inadequate response to therapy, poor compliance, or concerns regarding the increased potential for side effects at higher doses.

On the other hand, a drug called Ezetimibe (Zetia) works in the digestive tract. It is unique in the way it helps block absorption of cholesterol that comes from food. Ezetimibe (EZE) can complement statins in targeting both sources of cholesterol. EZE can be given as monotherapy to lower cholesterol levels in patients who are intolerant to statin or in whom treatment with statin is not appropriate. EZE can also be used in combination with a statin in patients whose cholesterol levels remain elevated despite treatment with statin alone. It can be either co-administered with the statin dose or given as a fixed-dose combination tablet (known as Vytorin) containing simvastatin and EZE. A combination tablet of atorvastatin and EZE (known as Liptruzet) and composite packs of rosuvastatin and EZE (known as Rosuzet) are also available around the world. The data for direct comparisons of these combination therapies for their effectiveness are very limited. The clinical trials for head-to-head comparisons of these combination therapies are very limited prompting for the indirect comparisons using network meta-analysis. This is the motivation and objective of our work in this article. The effects of statins in combination with EZE on LDL-C could vary due to possible drug interactions and need to be studied and is a subject of this article. The effects of statins have been well-studied and known in the literature. We estimate these effects through both direct

and indirect comparisons here in a network meta-analysis framework. Advantages of using combination therapy include greater efficacy through differing mechanisms of action, lower doses of individual drugs, and potential amelioration of side effects experienced with high doses of single agents (Morrone and others, 2012).

In network meta-analysis, more than two treatments are compared in different randomized pairwise or multi-arm trials (Lu and Ades, 2006). Approximately a quarter of randomized trials include more than two arms (Chan and Altman, 2005), while the presence of multi-arm trials adds complexity to the analysis. Moreover, in the evidence synthesis process, it is not rare to encounter treatments which are involved in very few trials or even only in one trial (Lu and Ades, 2006; Hong and others, 2013, 2016; Gwon and others, 2016). In the case, where some treatments are involved in only one trial, excluding such treatments sometimes can have important effects on the network meta-analysis results (Mills and others, 2013). Bayesian approaches are becoming more popular due to their flexibility and interpretability (Higgins and Whitehead, 1996; Lu and Ades, 2004, 2006). Meanwhile, random effects models are increasingly popular as a useful tool for network meta-analysis (Lu and Ades, 2006; Hong and others, 2013). Compared with fixed effects models, one of the advantages of random effects models is to allow for borrowing of strength from different trials. If all treatments are involved in multiple trials, it is desirable to allow the between-trial variances of the random effects to vary across treatments (arm-based model) or treatment comparisons (contrast-based model) (Lumley, 2002; Hong and others, 2016). It is a common practice to assume homogeneity of between-trial variations for all arms. The performance of homogeneous and heterogeneous variance models have been examined in Lu and Ades (2004, 2006, 2009). When the number of treatments is large and some treatments are involved only in a single trial, the variance parameters of the random effects in heterogeneous variance models cannot be estimated. In addition, an equal-correlation structure is often assumed for the correlation matrix of the random effects, and a value of half is usually assumed for the correlation (Lu and Ades, 2004, 2006). However, the equal-correlation structure is quite restrictive, since it does not allow for different correlations as well as arbitrary negative correlations among the random effects.

In this article, we consider a network meta-data consisting of 29 trials, 11 treatment arms (10 active treatments plus placebo), and 10 aggregate covariates. We first develop arm-based meta-regression models with multivariate random effects in order to compare these treatments while adjusting for aggregate covariates. A challenge that arises in dealing with 11-dimensional random effects is that some variances of the random effects cannot be estimated due to the fact that five treatments are involved only in a single trial. To circumvent this issue, we develop a general framework to formulate possible sets of the groups of treatments according to their clinical mechanisms of action, and further use the deviance information criterion (DIC) and the logarithm of the pseudo marginal likelihood (LPML) to select the best group. Unlike assuming known correlations, we specify a non-informative uniform prior for the correlation matrix and then develop a new localized Metropolis algorithm to generate the correlation matrix from its full conditional distribution via partial correlations. Finally, we carry out a detailed analysis of the network meta-data using the proposed methodology and our results in the treatment comparisons are generally consistent with direct comparisons reported in the literature.

The rest of this article is organized as follows. In Section 2, we provide a detailed description of the LDL-C network meta-data from 29 clinical trials, which motivates the proposed methodology. In Section 3, we fully develop the network meta-regression model, specify a multivariate normal distribution for random effects, and propose a grouping methodology for the covariance matrix. The complete-data likelihood and the observed-data likelihood are also given in Section 3. We present priors and posteriors in Section 4.1, develop a Markov chain Monte Carlo (MCMC) sampling algorithm, including collapsed Gibbs sampling and localized metropolis sampling in Section 4.2, and derive Bayesian model comparison criteria in Section 4.3. Section 5 gives a detailed analysis of the LDL-C network meta-data discussed in Section 2. We conclude the article with some discussion in Section 6.

2. THE LDL-C NETWORK META-DATA

A systematic search for randomized clinical trials using statins was conducted online using Google Scholar. The search yielded 78 trials: 32 Merck Sharp and Dohme (MSD) sponsored and 46 non-MSD sponsored trials. From these trials, all second-line studies (i.e. studies with patients on statin at study entry) were excluded. From the remaining first-line trials (i.e. studies with patients who were drug-naïve or rendered drug-naïve by wash-out at study entry), those with missing information on the response variable (LDL-C mean percent change) or study covariates (listed below) were excluded. The inclusion-exclusion flow diagram of studies is given in Figure 1 (a). This left us with 29 double-blind, randomized, active, or placebo-controlled clinical trials on adult treatment-naïve patients with primary hypercholesterolemia (15 MSD-sponsored and 14 non-MSD sponsored). These trials were conducted between 2002 and 2010 and study durations ranged from 5 to 12 weeks. Some trials had longer durations with titration of doses but only the data prior to the first titration were used in the analysis. The primary goal of these clinical trials was to evaluate the LDL-C lowering effects of different statins or ezetimibe plus statins. The treatments used in these trials were placebo (PBO), simvastatin (S), atorvastatin (A), lovastatin (L), rosuvastatin (R), pravastatin (P), Ezetimibe (E), and the combinations of S and E (SE), A and E (AE), L and E (LE) and P and E (PE). Ezetimibe is available at only one dose of 10 mg while the statins are available at multiple doses. In this article, the LDL-C lowering effects of different doses of each statin are combined to form the treatment group.

A network diagram (for LDL-C mean percent change) of the included treatments based on these 29 trials is presented in Figure 1 (b). Taking SE as an example, we can see from the diagram that (i) SE was included in trials 1, 3, 4, 5, 6, 7, 8, 9, 12, 13, 14, and 15, and 6596 patients were treated with SE in all these trials; (ii) SE was compared head-to-head with PBO, A, R, S, and E, while it is not compared head-to-head with AE, L, LE, PE, and P; and (iii) SE and PBO were compared head-to-head in trials 1, 3, and 6. Note that the size of the node is proportional to the total sample size of the corresponding treatment, and the width of the edge is related to the number of trials which include the direct comparisons.

The covariates considered here are baseline LDL-C (bl_ldlc) mean, baseline HDL-C (bl_hdlc) mean, baseline TG (bl_trig) mean, age mean, race proportion (of white), gender proportion (of male), body mass index (BMI) mean, proportion of medium statin potency, proportion of high statin potency, and trial duration. We consider mean percent change from baseline in LDL-C as the outcome variable. A summary of the covariates and outcome variables is given in Tables S1a and S1b of the [supplementary materials](#) available at *Biostatistics* online. Tables S2a and S2b of [supplementary material](#) available at *Biostatistics* online provide the title, treatment groups, treatment duration, and citation for the published primary manuscript for each trial. The patient entry criteria for these studies included in this meta-analysis are given in Tables S3a and S3b of [supplementary material](#) available at *Biostatistics* online. All head-to-head comparisons from these trials can be easily seen through Table S4 of the [supplementary material](#) available at *Biostatistics* online.

3. NETWORK META-REGRESSION MODELS

Suppose, we consider K randomized trials and a set of treatments $\mathcal{T} = \{0, 1, \dots, T\}$ from all K trials. The k th trial has $T^{(k)}$ treatments, which are denoted by $\mathcal{T}^{(k)} = \{t_1^{(k)}, \dots, t_{T^{(k)}}^{(k)}; t_\ell^{(k)} \in \mathcal{T}, \ell = 1, \dots, T^{(k)}\}$. Let $y_{t_\ell^{(k)}}^{(k)}$ denote the aggregate response, which generally represents the sample mean, and $S_{t_\ell^{(k)}}^{2(k)}$ denote the sample variance for $\ell = 1, 2, \dots, T^{(k)}$ and $k = 1, 2, \dots, K$. For the LDL-C network meta-data discussed in Section 2, we have $K = 29$ trials and $T = 10$ active treatment arms. We use 0 to denote PBO and 1–10 to denote S, A, L, R, P, E, SE, AE, LE, and PE, respectively. The first trial $k = 1$ includes treatments PBO, S, E, and SE, thus $T^{(1)} = 4$ and $\mathcal{T}^{(1)} = \{t_1^{(1)} = 0, t_2^{(1)} = 1, t_3^{(1)} = 6, t_4^{(1)} = 7\}$, which is a subset of $\mathcal{T}$.

Fig. 1. (a) Flow diagram for trials included in the network meta-analysis. (b) LDL-C network metadata diagram. Each node represents a treatment in the network metadata. The number associated with the treatment gives the total number of patients across all trials. Each edge represents the direct evidence comparing the treatments it connects, and the involved trial IDs are given for each head-to-head comparison.

Following [Yao and others \(2011\)](#), [Yao and others \(2015\)](#), and [Hong and others \(2017\)](#), we propose the following random effects model for the network meta-analysis

$$y_{t_\ell}^{(k)} = (\mathbf{x}_{t_\ell}^{(k)})^T \boldsymbol{\beta} + \gamma_{t_\ell}^{(k)} + \epsilon_{t_\ell}^{(k)}, \quad \epsilon_{t_\ell}^{(k)} \sim N\left(0, \frac{\sigma_{t_\ell}^{2(k)}}{n_{t_\ell}^{(k)}}\right), \quad (3.1)$$

and

$$\frac{(n_{t_\ell}^{(k)} - 1)S_{t_\ell}^{2(k)}}{\sigma_{t_\ell}^{2(k)}} \sim \chi_{(n_{t_\ell}^{(k)} - 1)}^2, \quad (3.2)$$

where $y_{t_\ell}^{(k)}$ and $S_{t_\ell}^{2(k)}$ are independent. The p -dimensional vector $\mathbf{x}_{t_\ell}^{(k)}$ represents the aggregate (arm-level) covariates and $\boldsymbol{\beta}$ is a p -dimensional vector of regression coefficients corresponding to the aggregate covariate vector $\mathbf{x}_{t_\ell}^{(k)}$.

The random effect of the $t_\ell^{(k)}$ th treatment within the k th trial, $\gamma_{t_\ell}^{(k)}$, is assumed to be independent of $\epsilon_{t_\ell}^{(k)}$. It captures the dependence of the $y_{t_\ell}^{(k)}$'s within the trial as well as the heterogeneity across trials. Let $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_T)'$ represent the $(T + 1)$ -dimensional vector of the overall treatment effects and Ω denote the $(T + 1) \times (T + 1)$ unknown covariance matrix. We define a collection of unit vectors $E^{(k)} = (e_{t_1}^{(k)}, e_{t_2}^{(k)}, \dots, e_{t_{T(k)}}^{(k)})$, where $e_{t_\ell}^{(k)} = (0, \dots, 1, \dots, 0)'$, $\ell = 1, \dots, T^{(k)}$ with $t_\ell^{(k)}$ th element equal to 1 and 0 otherwise. Thus, $E^{(k)}$ is a $(T + 1) \times T^{(k)}$ matrix. Also, let $(E^{(k)})^C$ be a $(T + 1) \times (T + 1 - T^{(k)})$ matrix, which consists of the columns of the $(T + 1) \times (T + 1)$ identity matrix I_{T+1} that are not included in $E^{(k)}$. We let $\boldsymbol{\gamma}^{(k)} = (\gamma_0^{(k)}, \gamma_1^{(k)}, \dots, \gamma_T^{(k)})'$ denote the vector of the $(T + 1)$ -dimensional random effects, which would be observed in the k th trial. Then, $\boldsymbol{\gamma}_o^{(k)} = (E^{(k)})^T \boldsymbol{\gamma}^{(k)}$ is the vector of the $T^{(k)}$ -dimensional random effects of the treatments that are actually observed in the k th trial while $\boldsymbol{\gamma}_m^{(k)} = ((E^{(k)})^C)^T \boldsymbol{\gamma}^{(k)}$ is the vector of the $(T + 1 - T^{(k)})$ -dimensional random effects of the treatments that are not included in the k th

trial. As an illustration, $(E^{(1)})^T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$ and $\boldsymbol{\gamma}_o^{(1)} = (E^{(1)})^T \boldsymbol{\gamma}^{(1)} = (\gamma_0^{(1)}, \gamma_1^{(1)}, \gamma_6^{(1)}, \gamma_7^{(1)})^T$. A detailed summary of the key notations is provided in Tables S4a and S4b of [supplementary material](#) available at *Biostatistics* online.

For the random effects, we assume $\boldsymbol{\gamma}^{(k)} \sim N_{T+1}(\boldsymbol{\gamma}, \Omega)$, which is a multivariate normal distribution with a $(T + 1)$ -dimensional vector of overall effects $\boldsymbol{\gamma}$ and a $(T + 1) \times (T + 1)$ positive definite covariance matrix Ω . Further, let $(\boldsymbol{\gamma}^{(k)})^R = \boldsymbol{\gamma}^{(k)} - \boldsymbol{\gamma}$, so that $(\boldsymbol{\gamma}^{(k)})^R \sim N_{T+1}(\mathbf{0}, \Omega)$. It is easy to show that $(\boldsymbol{\gamma}_o^{(k)})^R \sim N_{T^{(k)}}(\mathbf{0}, (E^{(k)})^T \Omega E^{(k)})$ and $(\boldsymbol{\gamma}_m^{(k)})^R \sim N_{T+1-T^{(k)}}(\mathbf{0}, ((E^{(k)})^C)^T \Omega (E^{(k)})^C)$, where $(\boldsymbol{\gamma}_o^{(k)})^R = (E^{(k)})^T (\boldsymbol{\gamma}^{(k)} - \boldsymbol{\gamma})$ and $(\boldsymbol{\gamma}_m^{(k)})^R = ((E^{(k)})^C)^T (\boldsymbol{\gamma}^{(k)} - \boldsymbol{\gamma})$, $k = 1, 2, \dots, K$.

Now, let $\boldsymbol{\epsilon}^{(k)} = \left(\epsilon_{t_1}^{(k)}, \epsilon_{t_2}^{(k)}, \dots, \epsilon_{t_{T(k)}}^{(k)} \right)$ denote the vector of random errors for the k th trial. Then $\boldsymbol{\epsilon}^{(k)} \sim N_{T^{(k)}}(\mathbf{0}, \Sigma^{(k)})$, where $\Sigma^{(k)} = \text{diag} \left(\frac{\sigma_{t_1}^{2(k)}}{n_{t_1}^{(k)}}, \frac{\sigma_{t_2}^{2(k)}}{n_{t_2}^{(k)}}, \dots, \frac{\sigma_{t_{T(k)}}^{2(k)}}{n_{t_{T(k)}}^{(k)}} \right)$ is a $T^{(k)} \times T^{(k)}$ diagonal matrix. Let

$\mathbf{X}^{(k)} = \begin{pmatrix} \mathbf{x}_{t_1^{(k)}}^{(k)} & \mathbf{x}_{t_2^{(k)}}^{(k)} & \dots & \mathbf{x}_{t_{T^{(k)}}}^{(k)} \end{pmatrix}'$ denote the $T^{(k)} \times p$ covariate matrix for the k th trial. Write $\mathbf{W}^{(k)} = (\mathbf{X}^{(k)} (E^{(k)})^T)$ and $\boldsymbol{\beta}^* = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$. Then the vector form of (3.1) is given by

$$\mathbf{y}^{(k)} = \mathbf{W}^{(k)} \boldsymbol{\beta}^* + (\boldsymbol{\gamma}_o^{(k)})^R + \boldsymbol{\epsilon}^{(k)}, \quad (3.3)$$

where $\mathbf{y}^{(k)} = \begin{pmatrix} y_{t_1^{(k)}}^{(k)} & y_{t_2^{(k)}}^{(k)} & \dots & y_{t_{T^{(k)}}}^{(k)} \end{pmatrix}'$.

Let $D_{oy} = \left\{ \left(y_{t_\ell}^{(k)}, n_{t_\ell}^{(k)}, \mathbf{x}_{t_\ell}^{(k)} \right), \ell = 1, 2, \dots, T^{(k)}, k = 1, 2, \dots, K \right\}$ and $D_{os} = \left\{ S_{t_\ell}^{2(k)}, n_{t_\ell}^{(k)}, \ell = 1, 2, \dots, T^{(k)}, k = 1, 2, \dots, K \right\}$. Then $D_o = D_{oy} \cup D_{os}$ denotes the observed data. Further, let $D_c = D_o \cup \left\{ \left(\gamma_{t_\ell}^{(k)} \right)^R, \ell = 1, 2, \dots, T^{(k)}, k = 1, 2, \dots, K \right\}$ denote the complete data. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^*, \Omega, \Sigma^*)$

denote the collection of all model parameters, where $\Sigma^* = \left(\sigma_{t_1^{(1)}}^{2(1)}, \dots, \sigma_{t_{T^{(1)}}}^{2(1)}, \dots, \sigma_{t_1^{(K)}}^{2(K)}, \dots, \sigma_{t_{T^{(K)}}}^{2(K)} \right)$.

Using (3.2) and (3.3) and the independence of $y_{t_\ell}^{(k)}$ and $S_{t_\ell}^{2(k)}$, the complete data likelihood function can be written as

$$\begin{aligned} L(\boldsymbol{\theta} \mid D_c) &= \prod_{k=1}^K \left((2\pi)^{-\frac{T^{(k)}}{2}} |\Sigma^{(k)}|^{-\frac{1}{2}} \exp \left\{ -\frac{(\mathbf{y}^{(k)} - \mathbf{W}^{(k)} \boldsymbol{\beta}^* - (\boldsymbol{\gamma}_o^{(k)})^R)^T (\Sigma^{(k)})^{-1} (\mathbf{y}^{(k)} - \mathbf{W}^{(k)} \boldsymbol{\beta}^* - (\boldsymbol{\gamma}_o^{(k)})^R)}{2} \right\} \right. \\ &\quad \times \prod_{\ell=1}^{T^{(k)}} \left[\frac{\left((n_{t_\ell}^{(k)} - 1) S_{t_\ell}^{2(k)} \right)^{\frac{n_{t_\ell}^{(k)} - 1}{2} - 1}}{\left(2\sigma_{t_\ell}^{2(k)} \right)^{\frac{n_{t_\ell}^{(k)} - 1}{2}} \Gamma \left(\frac{n_{t_\ell}^{(k)} - 1}{2} \right)} \exp \left\{ -\frac{(n_{t_\ell}^{(k)} - 1) S_{t_\ell}^{2(k)}}{2\sigma_{t_\ell}^{2(k)}} \right\} \times f((\boldsymbol{\gamma}^{(k)})^R \mid \Omega) \right] \Bigg), \end{aligned} \quad (3.4)$$

where $f((\boldsymbol{\gamma}^{(k)})^R \mid \Omega)$ is the probability density function corresponding to a $N_{T+1}(\mathbf{0}, \Omega)$ distribution. The random effects $(\boldsymbol{\gamma}^{(k)})^R$ can be directly integrated out from (3.4) because they follow a multivariate normal distribution. In equation 3.3, the random effects $(\boldsymbol{\gamma}_o^{(k)})^R$ are distributed as multivariate normal, and are independent of $\boldsymbol{\epsilon}^{(k)}$. Hence, we have $\mathbf{y}^{(k)} \mid \boldsymbol{\theta} \sim N(\mathbf{W}^{(k)} \boldsymbol{\beta}^*, (E^{(k)})' \Omega E^{(k)} + \Sigma^{(k)})$.

Therefore, the observed data likelihood function is given by

$$L(\boldsymbol{\theta} \mid D_o) = L(\boldsymbol{\theta} \mid D_{oy}) L(\Sigma^* \mid D_{os}), \quad (3.5)$$

$$\text{where } L(\boldsymbol{\theta} \mid D_{oy}) = \prod_{k=1}^K \left[(2\pi)^{-\frac{T(k)}{2}} |(E^{(k)})^T \Omega E^{(k)} + \Sigma^{(k)}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{y}^{(k)} - W^{(k)} \boldsymbol{\beta}^*)^T ((E^{(k)})^T \Omega E^{(k)} + \right. \right. \\ \left. \left. \Sigma^{(k)})^{-1} (\mathbf{y}^{(k)} - W^{(k)} \boldsymbol{\beta}^*) \right\} \right] \text{ and } L(\Sigma^* \mid D_{os}) = \prod_{k=1}^K \prod_{\ell=1}^{T(k)} \left[\frac{\left(\frac{(n_{t_\ell}^{(k)} - 1) S_{t_\ell}^{2(k)}}{n_{t_\ell}^{(k)} - 1} \right)^{\frac{n_{t_\ell}^{(k)} - 1}{2} - 1}}{(2\sigma_{t_\ell}^{2(k)})^{\frac{n_{t_\ell}^{(k)} - 1}{2}} \Gamma(\frac{n_{t_\ell}^{(k)} - 1}{2})} \exp \left\{ -\frac{(n_{t_\ell}^{(k)} - 1) S_{t_\ell}^{2(k)}}{2\sigma_{t_\ell}^{2(k)}} \right\} \right].$$

One of the major challenges for the proposed network meta-regression model is that only part of the covariance matrix Ω can be estimated due to the fact that some of the treatments are only included in a single trial. For example, treatments P, PE, L, LE, and AE were only included in trials 2, 10, and 11, respectively, for the LDL-C network meta-data. This makes estimation of the variances of the random effects corresponding to these treatments impossible. To overcome this problem, we assume that (i) these $T + 1$ treatments can be divided into G groups; and (ii) the random effects for those treatments within the same group share the same variance. Mathematically, we assume $\mathcal{T} = \bigcup_{g=1}^G \mathcal{G}_g$ such that $\mathcal{G}_g \cap \mathcal{G}_{g'} = \emptyset$ for all $g \neq g'$ and let $\{\tau_g^2, g = 1, 2, \dots, G\}$ denote the G distinct variances of the random effects. Then, we assume

$$\Omega_{jj} = \tau_g^2 \text{ for } j \in \mathcal{G}_g, \quad (3.6)$$

for $g = 1, 2, \dots, G$. We write the covariance matrix as $\Omega = V^{\frac{1}{2}} \boldsymbol{\rho} V^{\frac{1}{2}}$, where $V = \text{diag}(\Omega_{00}, \Omega_{11}, \dots, \Omega_{TT})$ is the matrix of the diagonal elements of Ω and $\boldsymbol{\rho}$ is the corresponding correlation matrix. Thus, the grouping of variances does not imply any grouping of correlations.

The determination of G and $\mathcal{G}_g$ is not an easy task. Here, we propose a two-step grouping strategy: (i) formulating possible sets of the groups of treatments based on their clinical mechanisms of action and (ii) using Bayesian model comparison criteria to select the best set of groups. For the LDL-C network meta-data, the random effect corresponding to the placebo arm is expected to have a smaller variance than the random effects corresponding to the active treatment arms since the placebo effect should be very similar across trials. Therefore, the placebo arm should stand alone as a group. In addition, statins (inhibit cholesterol synthesis in the liver) have a very different mechanism of action from EZE (inhibit the absorption of cholesterol by the small intestine). Therefore, the treatments with statins alone should not be classified into the same group as those with EZE arms. According to this strategy, we first formulate eight sets of G and $\mathcal{G}_g$ and then determine the best set of G and $\mathcal{G}_g$ according to Bayesian model comparison criteria.

4. BAYESIAN INFERENCE

4.1. Priors and posteriors

We assume that $\boldsymbol{\beta}^*$, Ω , and $\Sigma^{(k)}$, $k = 1, 2, \dots, K$ are independent a priori. We further assume $\boldsymbol{\beta}^* \sim$

$N_{p+T+1}(0, c_{01} I_{p+T+1})$. For $\Sigma^{(k)} = \text{diag} \left(\frac{\sigma_{t_1}^{2(k)}}{n_{t_1}^{(k)}}, \frac{\sigma_{t_2}^{2(k)}}{n_{t_2}^{(k)}}, \dots, \frac{\sigma_{T(k)}^{2(k)}}{n_{T(k)}^{(k)}} \right)$, we assume that $\sigma_{t_\ell}^{2(k)} \sim \text{IG}(a_{00}, b_{00})$,

$\ell = 1, 2, \dots, T^{(k)}$, $k = 1, 2, \dots, K$, that is, $p \left(\sigma_{t_\ell}^{2(k)} \mid a_{00}, b_{00} \right) \propto \left(\sigma_{t_\ell}^{2(k)} \right)^{-a_{00}-1} \exp \left\{ -\frac{b_{00}}{\sigma_{t_\ell}^{2(k)}} \right\}$. Write

$\boldsymbol{\rho} = (\rho_{ij})_{0 \leq i, j \leq T}$ and assume that $\pi((\rho_{01}, \rho_{12}, \rho_{02}, \rho_{23}, \rho_{13}, \dots, \rho_{0T})) \propto 1$ such that $\boldsymbol{\rho}$ is positive definite. We further assume $\Omega_{jj} = \tau_g^2 \sim IG(a_{0g}, b_{0g})$, for $j \in \mathcal{G}_g$, $g = 1, 2, \dots, G$.

Note that c_{01} , a_{00} , b_{00} , a_{0g} and b_{0g} are prespecified hyperparameters. In this article, we use $c_{01} = 100,000$, $a_{00} = 0.0001$, $b_{00} = 0.0001$, $a_{0g} = 0.01$, and $b_{0g} = 0.01$, $g = 1, 2, \dots, G$. We use the sampling algorithm based on partial correlations in Yao *and others* (2015) to sample $\boldsymbol{\rho}$ to ensure that $\boldsymbol{\rho}$ is a positive definite correlation matrix. Let $\boldsymbol{y}_o^R = ((\boldsymbol{y}_o^{(1)})^R)^T, ((\boldsymbol{y}_o^{(2)})^R)^T, \dots, ((\boldsymbol{y}_o^{(K)})^R)^T)^T$. From previous discussion and (3.4), the augmented posterior distribution is given by

$$\begin{aligned} & \pi(\boldsymbol{\beta}^*, \Omega, \Sigma^*, \boldsymbol{y}_o^R \mid D_o) \\ & \propto \prod_{k=1}^K |\Sigma^{(k)}|^{-\frac{1}{2}} \exp \left\{ -\frac{(\boldsymbol{y}^{(k)} - W^{(k)} \boldsymbol{\beta}^* - (\boldsymbol{y}_o^{(k)})^R)^T (\Sigma^{(k)})^{-1} (\boldsymbol{y}^{(k)} - W^{(k)} \boldsymbol{\beta}^* - (\boldsymbol{y}_o^{(k)})^R)}{2} \right\} L(\Sigma^* \mid D_{os}) \\ & \times \prod_{k=1}^K |(E^{(k)})' \Omega E^{(k)}|^{-\frac{1}{2}} \exp \left\{ -\frac{((\boldsymbol{y}_o^{(k)})^R)^T ((E^{(k)})' \Omega E^{(k)})^{-1} (\boldsymbol{y}_o^{(k)})^R}{2} \right\} \exp \left\{ -\frac{(\boldsymbol{\beta}^*)' \boldsymbol{\beta}^*}{2c_{01}} \right\} \\ & \times \prod_{k=1}^K \prod_{l=1}^{T^{(k)}} (\sigma_{t_\ell^{(k)}}^{2(k)})^{-a_{00}-1} \exp \left\{ -\frac{b_{00}}{\sigma_{t_\ell^{(k)}}^{2(k)}} \right\} \prod_{g=1}^G (\tau_g^2)^{-a_{0g}-1} \exp \left\{ -\frac{b_{0g}}{\tau_g^2} \right\}, \end{aligned} \quad (4.1)$$

where $L(\Sigma^* \mid D_{os})$ is defined in (3.5).

4.2. Computational development

The analytical evaluation of the posterior distribution of $\boldsymbol{\theta} = (\boldsymbol{\beta}^*, \Omega, \Sigma^*)$ given in (4.1) is not available. However, we can develop a MCMC sampling algorithm to sample from (4.1). The algorithm requires sampling the following parameters in turn from their respective full conditional distributions: (i) $[\Sigma^* \mid \boldsymbol{\beta}^*, \boldsymbol{y}_o^R, \Omega, D_o]$ and (ii) $[\boldsymbol{\beta}^*, \Omega, \boldsymbol{y}_o^R \mid \Sigma^*, D_o]$. For (ii), we use the modified collapsed Gibbs sampling technique in Chen *and others* (2000) via the identity $[\boldsymbol{\beta}^*, \Omega, \boldsymbol{y}_o^R \mid \Sigma^*, D_o] = [\boldsymbol{\beta}^*, \Omega \mid \Sigma^*, D_o][\boldsymbol{y}_o^R \mid \Sigma^*, \boldsymbol{\beta}^*, \Omega, D_o]$, and further $[\boldsymbol{\beta}^*, \Omega \mid \Sigma^*, D_o]$ is sampled in turn from the following full conditional distributions: (iia) $[\boldsymbol{\beta}^* \mid \Sigma^*, \Omega, D_o]$; (iib) $[\Omega \mid \Sigma^*, \boldsymbol{\beta}^*, D_o]$. For (iib), the sampling scheme is not straightforward. Following Section 4.1, Ω can be written as $V^{\frac{1}{2}} \boldsymbol{\rho} V^{\frac{1}{2}}$. The sampling process is thus divided into two parts, $V \mid \Sigma^*, \boldsymbol{\beta}^*, \boldsymbol{\rho}, D_o$ and $\boldsymbol{\rho} \mid \Sigma^*, \boldsymbol{\beta}^*, V, D_o$. For (ii), the modified collapsed Gibbs algorithm further requires sampling from (iic) $[\boldsymbol{y}_o^R \mid \Sigma^*, \boldsymbol{\beta}^*, \Omega, D_o]$.

The full conditional distributions for (i), (iia), and (iic), which are given in Appendix A of [supplementary material](#) available at *Biostatistics* online, are either an inverse gamma distribution or a multivariate normal distribution. Thus, sampling from $[\Sigma^* \mid \boldsymbol{\beta}^*, \boldsymbol{y}_o^R, \Omega, D_o]$, $[\boldsymbol{\beta}^* \mid \Sigma^*, \Omega, D_o]$, and $[\boldsymbol{y}_o^R \mid \Sigma^*, \boldsymbol{\beta}^*, \Omega, D_o]$ is straightforward. The full conditional distributions $[V \mid \Sigma^*, \boldsymbol{\beta}^*, \boldsymbol{\rho}, D_o]$ and $[\boldsymbol{\rho} \mid \Sigma^*, \boldsymbol{\beta}^*, V, D_o]$ are also given in Appendix A of [supplementary material](#) available at *Biostatistics* online. Sampling from these two conditional distributions is not trivial. In Appendix A of [supplementary material](#) available at *Biostatistics* online, we develop a modified localized Metropolis algorithm to sample from each of these two conditional distribution. The previous localized Metropolis algorithms used in Chen *and others* (2000) and Yao *and others* (2015) require the first and second derivatives of the logarithms of the full conditional densities. Unfortunately, computing the first and second derivatives of the log full conditional densities for $[V \mid \Sigma^*, \boldsymbol{\beta}^*, \boldsymbol{\rho}, D_o]$ and $[\boldsymbol{\rho} \mid \Sigma^*, \boldsymbol{\beta}^*, V, D_o]$ are prohibitive. The modified localized Metropolis algorithm avoids the direct computation of the second derivatives of these log full conditional densities.

4.3. Bayesian model comparison

Notice that in Section 3, we introduced the grouping approach to solve the partial estimability problem of the covariance matrix Ω . The grouping of the $T + 1$ variances into G groups motivates the idea of model comparison to select the appropriate grouping. We carry out model comparison using the DIC (Spiegelhalter and others, 2002) and the LPML (Ibrahim and others, 2001).

Using the previous notation, the collection of all model parameters is denoted by $\theta = (\beta^*, \Omega, \Sigma^*)$. Let $D_{oy}^{(k)} = \left\{ (y_{t_\ell}^{(k)}, n_{t_\ell}^{(k)}, \mathbf{x}_{t_\ell}^{(k)}), \ell = 1, \dots, T^{(k)} \right\}$ denote the response variables, sample sizes and covariates for the k_{th} trial. We define the trial-based deviance function $\text{Dev}^{(k)}(\theta)$ based on the observed-data likelihood corresponding to the response variables $\mathbf{y}^{(k)}$, that is $\text{Dev}^{(k)}(\theta) = -2 \log f(\theta | D_{oy}^{(k)})$, where $f(\theta | D_{oy}^{(k)})$ is the density function of a $N(W^{(k)}\beta^*, (E^{(k)})'\Omega E^{(k)} + \Sigma^{(k)})$ distribution. The trial-based DIC^(k) is given by $\text{DIC}^{(k)} = \text{Dev}^{(k)}(\bar{\theta}) + 2p_D^{(k)}$, where $\bar{\theta} = E[\theta | D_o]$ is the posterior mean of θ , $p_D^{(k)} = \overline{\text{Dev}^{(k)}(\theta)} - \text{Dev}^{(k)}(\bar{\theta})$ and $\overline{\text{Dev}^{(k)}(\theta)} = -2E_\theta[\log f(\theta | D_{oy}^{(k)})]$ is the posterior mean deviance. The overall deviance function $\text{Dev}(\theta)$ is thus given by $\text{Dev}(\theta) = -2 \log L(\theta | D_{oy}) = \sum_{k=1}^K \text{Dev}^{(k)}(\theta)$, and the DIC is

$$\text{DIC} = \text{Dev}(\bar{\theta}) + 2p_D = \sum_{k=1}^K \text{DIC}^{(k)}, \quad (4.2)$$

where $\text{Dev}(\bar{\theta}) = \sum_{k=1}^K \text{Dev}^{(k)}(\bar{\theta})$ is a measure of goodness of fit, and $p_D = \overline{\text{Dev}(\theta)} - \text{Dev}(\bar{\theta}) = \sum_{k=1}^K p_D^{(k)}$ is the effective number of model parameters.

To define the LPML, let $D_{oy}^{(-k)} = \left\{ (y_{t_\ell}^{(j)}, n_{t_\ell}^{(j)}, \mathbf{x}_{t_\ell}^{(j)}), \ell = 1, \dots, T_j, j = 1, \dots, k-1, k+1, \dots, K \right\}$ denote the response variables with the k_{th} trial deleted. The conditional predictive ordinate $\text{CPO}^{(k)}$ describes how much the response variables of the k_{th} trial supports the model and can be written as $\text{CPO}^{(k)} = \int f(\mathbf{y}^{(k)} | \theta) \pi(\theta | D_{oy}^{(-k)}, D_{os}) d\theta = \frac{1}{\int \frac{1}{f(\mathbf{y}^{(k)} | \theta)} \pi(\theta | D_{oy}^{(-k)}, D_{os}) d\theta}$, where $\pi(\theta | D_{oy}^{(-k)}, D_{os})$ is the posterior distribution of θ with data $(D_{oy}^{(-k)}, D_{os})$, and $f(\mathbf{y}^{(k)} | \theta)$ is the density function of a $N(W^{(k)}\beta^*, (E^{(k)})'\Omega E^{(k)} + \Sigma^{(k)})$ distribution.

The LPML can be used as a criterion-based measure for model selection, which is given by

$$\text{LPML} = \sum_{k=1}^K \log(\text{CPO}^{(k)}). \quad (4.3)$$

Since closed forms of $\text{CPO}^{(k)}$ are not available, we approximate them using Gibbs iterates $\{\theta^{(b)}, b = 1, \dots, B\}$. From [Dey and others \(1995\)](#), $\text{CPO}^{(k)}$ can be approximated using the Monte Carlo estimate $\widehat{\text{CPO}}^{(k)} = B \left[\sum_{b=1}^B \{f(\mathbf{y}^{(k)} | \theta^{(b)})\}^{-1} \right]^{-1}$.

5. ANALYSIS OF THE LDL-C NETWORK METADATA

Let $y_{t_\ell}^{(k)}$ be the mean percent change in LDL-C from the baseline value for the $t_\ell^{(k)}$ th treatment arm in the k th trial. The vector of covariates is $\mathbf{x}_{t_\ell}^{(k)} = (1, (\text{bl_ldlc})_{t_\ell}^{(k)}, (\text{bl_hdlc})_{t_\ell}^{(k)}, (\text{bl_tg})_{t_\ell}^{(k)}, \text{age}_{t_\ell}^{(k)}, \text{white}_{t_\ell}^{(k)}, \text{male}_{t_\ell}^{(k)}, \text{BMI}_{t_\ell}^{(k)}, \text{potency_med}_{t_\ell}^{(k)}, \text{potency_high}_{t_\ell}^{(k)}, \text{duration}_{t_\ell}^{(k)})^T$, and the corresponding regression coefficient vector is β .

Table 1. Description and comparison of models $\mathcal{M}_1 - \mathcal{M}_8$

Model	G	Ω	Description	DIC	LPML
$\mathcal{M}_1$	$G = 1$	$\Omega_{jj} = \tau_1^2, j = 0, \dots, 10$	The random effects for all 11 arms have the same variance	381.33	-165.30
$\mathcal{M}_2$	$G = 4$	$\Omega_{00} = \tau_1^2; \Omega_{jj} = \tau_2^2, j = 1, \dots, 5;$ $\Omega_{66} = \tau_3^2; \Omega_{jj} = \tau_4^2, j = 7, \dots, 10$	PBO alone; S/A/L/R/P; EZE alone; SE/AE/LE/PE	377.84	-164.68
$\mathcal{M}_3$	$G = 5$	$\Omega_{00} = \tau_1^2; \Omega_{11} = \tau_2^2;$ $\Omega_{jj} = \tau_3^2, j = 2, \dots, 5;$ $\Omega_{66} = \tau_4^2; \Omega_{jj} = \tau_5^2, j = 7, \dots, 10$	PBO alone; S alone; A/L/R/P; EZE alone; SE/AE/LE/PE	373.24	-161.93
$\mathcal{M}_4$	$G = 5$	$\Omega_{00} = \tau_1^2; \Omega_{44} = \tau_2^2;$ $\Omega_{jj} = \tau_3^2, j = 1, 2, 3, 5;$ $\Omega_{66} = \tau_4^2; \Omega_{jj} = \tau_5^2, j = 7, \dots, 10$	PBO alone; R alone; S/A/L/P; EZE alone; SE/AE/LE/PE	371.50	-161.87
$\mathcal{M}_5$	$G = 5$	$\Omega_{00} = \tau_1^2; \Omega_{22} = \tau_2^2;$ $\Omega_{jj} = \tau_3^2, j = 1, 3, 4, 5;$ $\Omega_{66} = \tau_4^2; \Omega_{jj} = \tau_5^2, j = 7, \dots, 10$	PBO alone; A alone; S/L/R/P; EZE alone; SE/AE/LE/PE	377.85	-164.90
$\mathcal{M}_6$	$G = 6$	$\Omega_{00} = \tau_1^2; \Omega_{11} = \tau_2^2;$ $\Omega_{44} = \tau_3^2; \Omega_{jj} = \tau_4^2, j = 2, 3, 5;$ $\Omega_{66} = \tau_5^2; \Omega_{jj} = \tau_6^2, j = 7, \dots, 10$	PBO alone; S alone; R alone; A/L/P; EZE alone; SE/AE/LE/PE	369.72	-161.19
$\mathcal{M}_7$	$G = 6$	$\Omega_{00} = \tau_1^2; \Omega_{11} = \tau_2^2;$ $\Omega_{22} = \tau_3^2; \Omega_{jj} = \tau_4^2, j = 3, 4, 5;$ $\Omega_{66} = \tau_5^2; \Omega_{jj} = \tau_6^2, j = 7, \dots, 10$	PBO alone; S alone; A alone; L/R/P; EZE alone; SE/AE/LE/PE	371.84	-161.64
$\mathcal{M}_8$	$G = 6$	$\Omega_{00} = \tau_1^2; \Omega_{22} = \tau_2^2;$ $\Omega_{44} = \tau_3^2; \Omega_{jj} = \tau_4^2, j = 1, 3, 5;$ $\Omega_{66} = \tau_5^2; \Omega_{jj} = \tau_6^2, j = 7, \dots, 10$	PBO alone; A alone; R alone; S/L/P; EZE alone; SE/AE/LE/PE	368.33	-160.50

Model $\mathcal{M}_1$ represents that all treatment arms share the same variance; model $\mathcal{M}_2$ has 4 groups, in which PBO alone is in the first group, all statins (S, A, L, R, P) are in the second group, EZE is in the third group, and all statins with EZE (SE, AE, LE, PE) are in the fourth group; models $\mathcal{M}_3 - \mathcal{M}_5$ are similar to model $\mathcal{M}_2$ but separate one statin (S, A or R), which all are involved in multiple trials, from the other statins; models $\mathcal{M}_6 - \mathcal{M}_8$ are also similar to model $\mathcal{M}_2$ but separate two statins (S and R, S and A, or A and R) from the other statins.

We fit the meta-regression model defined in (3.1) and (3.2) to the LDL-C network meta-data using the sampling algorithm proposed in Section 4.2. Let $\mathcal{M}_1, \dots, \mathcal{M}_8$ represent eight different groupings whose definitions are given in Table 1. Among the eight groupings, model $\mathcal{M}_1$ is the simplest model and therefore serves as the “benchmark” model. We computed the DICs defined in (4.2) and the LPMLs defined in (4.3) under models $\mathcal{M}_1 - \mathcal{M}_8$ and reported their results in Table 1. Since some treatments in trials 2, 10, and 11 are not included in any other trials, the calculation of the LPML excluded those trials. We see from Table 1 that (i) the DIC value under the 1 group model $\mathcal{M}_1$ (381.33) is larger than the DIC value under the 4 group model $\mathcal{M}_2$ (377.84), which implies that separating the variances of statins, EZE and statins with EZE from the variance of PBO is necessary for a better model fit; (ii) among the five group models $\mathcal{M}_3 - \mathcal{M}_5$, separating the variance of R from the all-statin group (i.e. $\mathcal{M}_4$) has the smallest DIC value (371.50), which is also smaller than the DIC value under model $\mathcal{M}_2$; and (iii) among the six group models $\mathcal{M}_6 - \mathcal{M}_8$, separating the variances of A and R from the all-statin group (i.e. $\mathcal{M}_8$) has the smallest DIC value (368.33), which is also smaller than

Fig. 2. Individual DIC and $\log(\text{CPO})$ plots of model $\mathcal{M}_1$ versus model $\mathcal{M}_8$. The plot of $\text{DIC}^{(k)}$'s is on the left and the plot of $\log(\text{CPO}^{(k)})$'s is on the right. The filled circle points represent the trials that favor the x -axis model compared to the y -axis model and the empty triangle points are *vice versa*. Differences of $\text{DIC}^{(k)}$'s or $\log(\text{CPO}^{(k)})$'s between two models larger than one are labeled with trial IDs.

the DIC value under model $\mathcal{M}_4$. The DIC values under models $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_4$, $\mathcal{M}_8$ indicate that the smallest DIC for each fixed G is a decreasing function of G and the “best” DIC value is attained at $G = 6$. The LPML values under $\mathcal{M}_1 - \mathcal{M}_8$ behave similarly to these DIC values and the LPML value under model $\mathcal{M}_8$ is the largest (-160.50), which is consistent with the smallest DIC value under model $\mathcal{M}_8$.

The trial-based $\text{DIC}^{(k)}$'s and $\text{CPO}^{(k)}$'s defined in Section 4.3 are also computed. Plots of the $\text{DIC}^{(k)}$'s and $\log(\text{CPO}^{(k)})$'s under model $\mathcal{M}_8$ versus model $\mathcal{M}_1$ are shown in Figure 2. In the $\text{DIC}^{(k)}$'s plot from Figure 2, 21 dots are red and 8 dots are blue. This implies that 21 out of 29 trials favor model $\mathcal{M}_8$ to model $\mathcal{M}_1$ in terms of individual DIC. Specifically, the values of DIC_1 , DIC_3 , DIC_6 , DIC_{10} , and DIC_{21} under model $\mathcal{M}_8$ are much smaller than those under model $\mathcal{M}_1$. In the $\log(\text{CPO}^{(k)})$'s plot from Figure 2, there are 22 out of 26 trials in favor of model $\mathcal{M}_8$. The value of $\log(\text{CPO}_{21})$ under model $\mathcal{M}_8$ are significantly larger than that under model $\mathcal{M}_1$. These results suggest that trials may favor different models, however, compared with the “benchmark” model $\mathcal{M}_1$, more trials favor model $\mathcal{M}_8$. This is consistent with the results in Table 1.

The posterior estimates, including posterior means, posterior standard deviations (SDs), and 95% highest posterior density (HPD) intervals of the parameters under model $\mathcal{M}_8$ and $\mathcal{M}_1$ are reported in Table 2 and Table S6 of [supplementary material](#) available at *Biostatistics* online. We see from these two tables that the posterior estimates for the overall treatment effects given the covariates were similar under the two models. Except for PBO, patients on all other treatments had substantial percent changes from baseline in LDL-C (i.e. the 95% HPD intervals did not contain 0). The estimate for γ_8 (i.e., AE) was the lowest (-51.93) which indicates that the treatment AE had the highest percent change from baseline in LDL-C. Among the ten covariates, the baseline HDL-C regression coefficient had an HPD interval not containing zero under model $\mathcal{M}_8$. The posterior mean (SD) and 95% HPD interval for β_2 (i.e. baseHDL-C) were -1.49 (0.74) and $(-2.93, -0.02)$. This result indicates that there was a substantial improvement in LDL-C from baseline for higher baseline value of HDL-C. In contrast, no covariates coefficients were

Table 2. Posterior estimates of the parameters under model $\mathcal{M}_8$

Variables	Parameter	Posterior		95% HPD Interval	Posterior $P(\gamma_j > 0)$
		Mean	SD		
baseLDLC	β_1	-0.10	0.63	(-1.30, 1.18)	
baseHDL	β_2	-1.49	0.74	(-2.93, -0.02)	
baseTG	β_3	0.52	0.60	(-0.63, 1.72)	
age	β_4	-0.74	0.54	(-1.79, 0.35)	
white	β_5	-0.79	0.53	(-1.87, 0.20)	
male	β_6	-1.21	0.75	(-2.64, 0.29)	
BMI	β_7	0.41	0.60	(-0.76, 1.58)	
potency_med	β_8	3.40	3.06	(-2.38, 9.65)	
potency_high	β_9	-1.69	3.40	(-8.19, 5.19)	
duration	β_{10}	1.09	0.63	(-0.19, 2.29)	
PBO	γ_0	1.20	5.47	(-9.69, 11.87)	0.59
S	γ_1	-40.43	1.11	(-42.63, -38.26)	0.00
A	γ_2	-44.60	1.64	(-47.81, -41.36)	0.00
L	γ_3	-27.84	3.89	(-35.35, -20.01)	5×10^{-5}
R	γ_4	-42.22	2.14	(-46.48, -38.08)	0.00
P	γ_5	-27.81	3.91	(-35.30, -19.96)	5×10^{-5}
E	γ_6	-18.99	5.41	(-29.49, -8.25)	1.35×10^{-3}
SE	γ_7	-47.48	2.19	(-51.72, -43.10)	0.00
AE	γ_8	-51.93	4.52	(-60.67, -42.74)	0.00
LE	γ_9	-47.84	4.55	(-56.81, -38.85)	0.00
PE	γ_{10}	-47.02	4.66	(-56.07, -37.65)	0.00
Variance of random effects	τ_1^2	2.14	3.05	(0.07, 6.95)	
	τ_2^2	4.90	2.91	(1.06, 10.73)	
	τ_3^2	14.65	7.15	(4.62, 28.43)	
	τ_4^2	2.43	3.82	(0.07, 7.61)	
	τ_5^2	2.21	3.58	(0.07, 7.33)	
	τ_6^2	15.24	8.57	(4.19, 31.14)	

significant under model $\mathcal{M}_1$. The posterior mean (SD) of τ_1^2 (i.e. the variance shared by the random effects for all 11 arms) under model $\mathcal{M}_1$ was 8.07 (1.83). Under the 6 group model $\mathcal{M}_8$, the posterior means (SDs) of τ_1^2 - τ_6^2 were 2.14 (3.05), 4.90 (2.91), 14.65 (7.15), 2.43 (3.82), 2.21 (3.58), and 15.24 (8.57), respectively. Note that under $\mathcal{M}_8$, the posterior mean of τ_1^2 (i.e. the variance of the random effect for PBO) is much smaller than the posterior mean of τ_1^2 under model $\mathcal{M}_1$, where the random effect for PBO was assumed to share the same variance with the random effects for other treatments. Moreover, the posterior estimates of the variances varied a lot between different groups under model $\mathcal{M}_8$. Therefore, the one group model $\mathcal{M}_1$ cannot fit the data well because the variances of the random effects were different among these treatments, which further confirmed the results of DICs and LPMLs in Table 1. The last column of Table 2 reported the posterior probability that each treatment effect was positive. Since the goal was to see reduction on LDL-C, the posterior probability $P(\gamma_j > 0)$ can be seen as an analogy to a P -value for each effect. A small $P(\gamma_j > 0)$ indicated strong evidence against the statement that the treatment was not effective.

Table S7 in Appendix B of [supplementary material](#) available at *Biostatistics* online presented the posterior means, posterior SDs, and 95% HPD intervals for the pairwise differences in treatments means (the mean percent reductions in LDL-C from the baseline) after adjusting for the aggregate covariates

under model $\mathcal{M}_8$. The following observations are noteworthy. First, as expected, all statins, EZE and statins with EZE combinations provided a substantially higher reduction in LDL-C than PBO. Second, among 10 pairwise comparisons between five statins, (a) S, A, and R provided a significantly higher LDL-C reduction than L and P, (b) A provided a significantly higher LDL-C reduction than S, and (c) the remaining three pairwise comparisons were not significant. Third, statins provided significantly higher LDL-C reduction than EZE. Fourth, considering the statins with EZE combination therapies, (a) the mean LDL-C reductions for these four combination therapies were not significantly different from each other, (b) all four combination therapies provided a significantly higher LDL-C reduction than corresponding statin mono-therapy and also E mono-therapy except with AE, in which case, the 95% HPD interval for the difference in LDL-C reductions between AE and A was $(-16.11, 1.41)$.

Our fitted model $\mathcal{M}_8$ did yield numerically a higher LDL-C reduction for AE than A. In the clinical literature, this difference in LDL-C reductions between AE and A is known to be highly significant through direct comparisons (Ballantyne and others, 2003). For comparison purposes, we also fitted model $\mathcal{M}_8$ without covariates. Table S8 in Appendix B of [supplementary material](#) available at *Biostatistics* online presented the pairwise comparisons which were not adjusted by aggregate covariates in model $\mathcal{M}_8$. The difference in LDL-C reductions due to AE and A became highly significant. The posterior mean (SD) of the difference was -15.97 (3.62) yielding $(-23.07, -8.74)$ and $(-32.53, -0.15)$ as 95% and 99.9% HPD intervals, respectively. Likewise, the difference in LDL-C reductions between R and S as well as between R and A became significant under model $\mathcal{M}_8$ without covariates (95% HPD interval, $[-13.04, -6.00]$ and $[-8.35, -4.38]$, respectively). The opposite was the case for the LDL-C reduction difference between A and S (95% HPD interval, $[-6.78, 0.43]$). All these three network meta-analysis results under model $\mathcal{M}_8$ without covariates were consistent with the direct comparison results from Insull and others (2007). In general, most estimates of the treatment differences from the network meta-analysis under model $\mathcal{M}_8$ were consistent with the direct comparisons (if there were any). Some of the treatment differences were in the right direction numerically but did not reach significance mainly due to the inclusion of covariates. Finally, the posterior estimates of the correlation parameters under model $\mathcal{M}_8$ are reported in Table S9 in Appendix B of [supplementary material](#) available at *Biostatistics* online. From this table, we see that none of the correlation parameters are significant since all of the 95% HPD intervals contain 0.

The absolute and cumulative probabilities under model $\mathcal{M}_8$ for all 11 treatments taking each possible rank were plotted in Figures 3 and S1 in Appendix B of [supplementary material](#) available at *Biostatistics* online. The probabilities in Figure 3 were adjusted by including the covariates while the probabilities in Figure S1 were not. Also reported was the surface under the cumulative ranking curve (SUCRA) (Salanti and others, 2011), which was used to estimate the ranking probabilities for all treatments to obtain a treatment hierarchy. SUCRA can also be used to calculate the normalized mean rank. The mean rank for a treatment is $1 + (1 - p)(T - 1)$, where p is the SUCRA for the treatment and T is the number of treatments. The order of treatments (in descending order) suggested by Figure 3 was AE, SE, LE, PE, A, R, S, L, P, E, and PBO. This suggested ranking order was consistent with the magnitudes of the estimated LDL-C percent reductions in Table 2. The order of treatments according to SUCRAs shown in Figure S1 was AE, SE, R, A, LE, PE, S, P, L, E, and PBO. Thus, the ranking of treatments was changed if the aggregate covariates were not included.

In all of the Bayesian computations, we used 20 000 MCMC samples, which were taken from every fifth iteration, after a burn-in of 5000 iterations for each model to compute all posterior estimates, including posterior means, posterior SDs, 95% HPD intervals, DICs and LPMLs, and cumulative ranking curves. The convergence of the MCMC sampling algorithm was checked using several diagnostic procedures discussed in Chen and others (2000). The HPD intervals were computed via the Monte Carlo method developed by Chen and Shao (1999).

Fig. 3. Plots of ranking probabilities for all treatment arms. The dashed line represents the absolute probability and the solid line represents the cumulative probability. SUCRA is the percentage of efficacy of a treatment on the outcome, which would be one when a treatment is certain to be the best and zero when a treatment is certain to be the worst.

6. DISCUSSION

Our methodology and analysis was motivated by the fact that the head-to-head clinical trials directly comparing all the treatments of interest is not possible in clinical research due to time, resources and other practical constraints. We have used real data from clinical trials on lipid lowering therapies to motivate our research and illustrate the network meta-analysis methodology developed here. Although, we do not know the ground truth, the results obtained for comparisons among LDL-C lowering therapies turn out to be fairly consistent with what is known in the clinical literature, and hence are supportive of the methodology and assumptions used here.

In Appendix A of [supplementary material](#) available at *Biostatistics* online, we use the modified localized Metropolis algorithm to generate a positive definite correlation matrix via the partial correlations (Joe, 2006). We also implement the modified localized Metropolis algorithm by reparameterizing the correlation matrix using a spherical co-ordinate system (Lu and Ades, 2009). Both reparameterization methods yield in the similar posterior estimates, as shown, for an example, in Table S10 of Appendix B of [supplementary material](#) available at *Biostatistics* online.

There are indeed two fundamental approaches to network meta-analysis in that one can use arm-based models or contrast-based models. The arguments for using a contrast-based model are strong in the case that there is a natural base treatment for each study for which to build the model. In our network, we do

not have a natural base treatment for some of the studies, and thus picking an arbitrary base treatment for some of the studies in the network may lead to biased assessments of the treatment effect and inappropriate inference. We believe that arm-based methods are most useful in settings in which there is not a natural base treatment for some studies, as is this the case here with our network. Thus, although we are aware of the debate (Dias and Ades, 2016; Hong and others, 2016) and are aware that contrast-based methods can handle confounding when there is a natural base treatment, arm-based methods may be more suitable in settings where no natural base treatment exists for many of the studies. It is for this reason, we have decided to take an arm-based modeling approach, and we believe that our arm-based approach is useful in our particular network meta-analysis setting.

As we have seen from this network meta-data, the dimension of the covariance matrix of the random effects is high and some treatments are included in very few trials or even in a single trial. Therefore, many correlation coefficients among the random effects cannot be estimated. Unlike the variances of the random effects, the correlation coefficients are bounded. Therefore, we simply assume a uniform prior for the correlation matrix ρ in our analysis. Due to the positive definite constraint of ρ , we have developed a Metropolis-within-Gibbs sampling algorithm to generate ρ via partial correlations. To allow for borrowing of strength from different pairs of correlation coefficients, a potential extension of the grouping approach for the variances of the random effects is to assume that some pairs of treatments share the same correlations. However, determining the number of groups and selecting group membership for correlation coefficients is much more challenging than for the variances of the random effects. From Table S1a, we see that the $S_{\ell}^{2(k)}$'s are generally large and the values of the $S_{\ell}^{(k)}$'s range from 10.50 to 18.04. Similar to the meta-regression model assumed for the mean response, a potential extension is to assume a log-linear meta regression model for $\sigma_{\ell}^{2(k)}$ in (3.1) and (3.2). These extensions are currently under investigation.

As an extension of the proposed grouping approach, both the total number of groups and group membership allocation are assumed to be random in the model; and then a reversible jump MCMC algorithm is developed to sample G and the membership allocation. This alternative approach may provide an empirical justification of the grouping based on the clinical relevance, which is certainly another promising future research project.

7. SOFTWARE

Computer code was written for the FORTRAN 95 compiler. We built an R package which includes the LDL-C network meta-data and serves as an interface calling the FORTRAN code within R. The R-package with the built-in data used in this article is available at <https://github.com/epochholly/Network-Meta-Analysis-Bayesian-Inference-Multivariate-Random-Effects>.

SUPPLEMENTARY MATERIAL

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

ACKNOWLEDGMENTS

We would like to thank the Editor, an Associate Editor, and two referees for their very helpful comments and suggestions, which have led to a much improved version of the article. *Conflict of Interest*: None declared.

FUNDING

National Institutes of Health (GM70335 and P01CA142538 to M.-H.C. and J.G.I.). Intramural Research Program of National Institutes of Health and National Cancer Institute (S.K.).

REFERENCES

- ADEDINSEWO, D., TAKA, N., AGASTHI, P., SACHDEVA, R., RUST, G. AND ONWUANYI, A. (2016). Prevalence and factors associated with statin use among a nationally representative sample of US adults: National Health and Nutrition Examination Survey, 2011–2012. *Clinical Cardiology* **9**, 491–496.
- BALLANTYNE, C. M., HOURI, J., NOTARBARTOLO, A., MELANI, L., LIPKA, L. J., SURESH, R., SUN, S., LEBEAUT, A. P., SAGER, P. T. AND VELTRI, E. P. (2003). Effect of ezetimibe coadministered with atorvastatin in 628 patients with primary hypercholesterolemia. *Circulation* **19**, 2409–2415.
- CHAN, A. W. AND ALTMAN, D. G. (2005). Epidemiology and reporting of randomised trials published in PubMed journals. *The Lancet* **9465**, 1159–1162.
- CHEN, M.-H. AND SHAO, Q. M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics* **1**, 69–92.
- CHEN, M.-H., SHAO, Q. M. AND IBRAHIM, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer.
- DEY, D. K., KUO, L. AND SAHU, S. K. (1995). A Bayesian predictive approach to determining the number of components in a mixture distribution. *Statistics and Computing* **4**, 297–305.
- DIAS, S. AND ADES, A. E. (2016). Absolute or relative effects? Arm-based synthesis of trial data. *Research Synthesis Methods* **7**, 23–28.
- EXPERT PANEL ON DETECTION, EVALUATION, AND TREATMENT OF HIGH BLOOD CHOLESTEROL IN ADULTS. (2001). Executive summary of the Third Report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III). *Journal of American Medical Association* **19**, 2486–2497.
- GWON, Y., MO, M., CHEN, M.-H., LI, J., XIA, H. A. AND IBRAHIM, J. G. (2016). Network meta-regression for ordinal outcomes: applications in comparing Crohn’s disease treatments. *Technical Report 16–28*, Department of Statistics, University of Connecticut.
- HONG, H., CARLIN, B. P., SHAMLIYAN, T. A., WYMAN, J. F., RAMAKRISHNAN, R., SAINFORT, F. AND KANE, R. L. (2013). Comparing Bayesian and frequentist approaches for multiple outcome mixed treatment comparisons. *Medical Decision Making* **5**, 702–714.
- HONG, H., CHU, H., ZHANG, J. AND CARLIN, B. P. (2016). A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. *Research Synthesis Methods* **7**, 6–22.
- HONG, H., PRICE, K. L., FU, H. AND CARLIN, B. P. (2017). Bayesian network meta-analysis for multiple endpoints. In: Gatsonis, C. and Morton, C. S. (editors), *Methods in Comparative Effectiveness Research*, Chapter 12. Taylor & Francis Group, pp. 385–407.
- HIGGINS, J. AND WHITEHEAD, A. (1996). Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* **24**, 2733–2749.
- IBRAHIM, J. G., CHEN, M.-H. AND SINHA, D. (2001). *Bayesian Survival Analysis*. New York: Springer.
- INSULL, W., GHALI, J. K., HASSMAN, D. R., YCAS, J. W., GANDHI, S. K. AND MILLER, E. (2007). Achieving low-density lipoprotein cholesterol goals in high-risk patients in managed care: comparison of rosuvastatin, atorvastatin, and simvastatin in the SOLAR trial. *Mayo Clinic Proceedings* **5**, 543–550.
- JOE, H. (2006). Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis* **10**, 2177–2189.

- LU, G. AND ADES, A. E. (2004). Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine* **20**, 3105–3124.
- LU, G. AND ADES, A. E. (2006). Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association* **474**, 447–459.
- LU, G. AND ADES, A. E. (2009). Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics* **10**, 792–805.
- LUMLEY, T. (2002). Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine* **16**, 2313–2324.
- MILLS, E. J., KANTERS, S., THORLUND, K., CHAIMANI, A., VERONIKI, A. A. AND IOANNIDIS, J. P. (2013). The effects of excluding treatments from network meta-analyses: survey. *British Medical Journal* **347**, f5195.
- MORRONE, D., WEINTRAUB, W. S., TOTH, P. P., HANSON, M. E., LOWE, R. S., LIN, J., SHAH, A. K. AND TERSHAKOVEC, A. M. (2012) Lipid-altering efficacy of ezetimibe plus statin and statin monotherapy and identification of factors associated with treatment response: A pooled analysis of over 21,000 subjects from 27 clinical trials. *Atherosclerosis* **223**, 251–261.
- NCHSTATS. (2013). *A Blog of the National Center for Health Center for Health Statistics*. <https://nchstats.com/2013/11/14/statistics-on-statin-use/>.
- ROSS, G. (2015). *Too Few Americans Take Statins, CDC Study Reveals*. <http://acsh.org/news/2015/12/04/cdc-study-reveals-that-too-few-americans-are-on-statins>.
- SALANTI, G., ADES, A. E. AND IOANNIDIS, J. P. (2011). Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *Journal of Clinical Epidemiology* **2**, 163–171.
- SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. AND VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **4**, 583–639.
- YAO, H., CHEN, M.-H. AND QIU, C. (2011). Bayesian modeling and inference for meta data with applications in efficacy evaluation of an allergic rhinitis drug. *Journal of Biopharmaceutical Statistics* **5**, 992–1005.
- YAO, H., KIM, S., CHEN, M.-H., IBRAHIM, J. G., SHAH, A. K. AND LIN, J. (2015). Bayesian inference for multivariate meta-regression with partially observed within-study sample covariance matrix. *Journal of the American Statistical Association* **510**, 528–544.

[Received February 24, 2017; revised March 5, 2018; accepted for publication March 18, 2018]