Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
2017, Vol. 26(6) 2885–2896

DOI: 10.1177/0962280215615003
journals.sagepub.com/home/smm
SAGE

A joint Bayesian approach for the analysis of response measured at a primary endpoint and longitudinal measurements

Zeynep Kalaylioglu[1] and Haydar Demirhan[2]

Abstract

Joint mixed modeling is an attractive approach for the analysis of a scalar response measured at a primary endpoint and longitudinal measurements on a covariate. In the standard Bayesian analysis of these models, measurement error variance and the variance/covariance of random effects are a priori modeled independently. The key point is that these variances cannot be assumed independent given the total variation in a response. This article presents a joint Bayesian analysis in which these variance terms are a priori modeled jointly. Simulations illustrate that analysis with multivariate variance prior in general lead to reduced bias (smaller relative bias) and improved efficiency (smaller interquartile range) in the posterior inference compared with the analysis with independent variance priors.

Keywords

Multivariate log gamma distribution, random effects, variance components, variance prior, longitudinal data

1 Introduction

In many researches, especially in health studies, interest lies in associating the response measured at the end of a period with the features of a longitudinal process throughout the period. Wang et al.[1] developed a joint modeling approach for the analysis of such data in which two models, one for the scalar response and the other one is for the longitudinal covariate, are joined together through a set of common latent factors. A generalized random effects model is used to capture the trend of longitudinal covariate measurements over time. Response model is a generalized linear model with fixed covariates and the subject-specific coefficients from the longitudinal model. Response model associates the cross-sectional response measured at a primary endpoint with the features of the longitudinal covariate process such as its slope and intercept. Efficient sufficiency and conditional score estimation for the joint model are developed by Li et al.[2] and Li et al.[3] for single and multiple longitudinal covariates, respectively, in which no assumption is required for the random effects. Li et al.[4] proposed a semiparametric extension to these models to be used when there is plausible distributional assumption about random effects. A standard Bayesian analysis of the joint model is given by Horrocks and van den Heuvel.[5] In this article, we focus on the Bayesian analysis of the model, variance priors in particular. In such models where total variation in a response is a *sum* of different variance components, we propose that it is natural to view this total variation as composite and its constituents, namely measurement error variances and random effects variances/covariances not independent and thus should be a priori modeled jointly. Our proposal goes beyond the current model of interest to a wide range of models and study designs in which different variance components are present.

Joint a priori modeling of variance parameters in a hierarchical model was first undertaken in the study of Demirhan and Kalaylioglu.[6] They considered a normal linear growth model with uncorrelated random coefficients and developed a strategy in which error variance and the variances of random effects were a priori modeled jointly. The rationale for considering a joint prior model for the variance parameters in a hierarchical model is the simple fact that the total variation in a response is composite and so there is a

[1]Department of Statistics, Middle East Technical University, Ankara, Türkiye
[2]Department of Statistics, Hacettepe University, Beytepe, Ankara, Türkiye

Corresponding author:
Zeynep Kalaylioglu, Department of Statistics, Middle East Technical University, 06800, Ankara, Türkiye.
Email: kzeynep@metu.edu.tr

certain intrinsic dependency between its constituents. In that approach, random effects variances (log transformed) and the error variance are stacked up into a vector and a multivariate prior distribution is assigned to account for the intrinsic dependency among them. For the multivariate prior distribution, in addition to some mainstream multivariate distributions such as multivariate normal and multivariate skew normal, they considered a fresh one namely generalized multivariate log gamma (G-MVLG) by Demirhan and Hamurkaroglu.[7] Simulation study therein, in which Bayesian analysis with joint prior model for the variance parameters and that with independent priors on error variance and random effects variances are compared, revealed that joint prior for variance components improved efficiency (smaller posterior variance) and diminished bias in Bayesian estimators of random coefficients and their population averages were compared to the the standard Bayesian analysis where independent variance priors are used. Simulation results also showed that, among the multivariate distributions considered therein, G-MVLG provided the smallest mean-squared errors in all the various different scenarios considered. Their study was limited to models with uncorrelated random effects and statistical and computational efficiency of the method for models with correlated random effects remained as a question.

Given the benefits of the approach in hierarchical models with uncorrelated random effects, the idea therein is now extended to the more common way of modeling in which random effects are correlated. The objectives of this article are to (i) develop a joint variance prior approach for models with correlated random effects, particularly in the context of joint model, which is used for the analysis of response measured at an endpoint and longitudinal covariates; (ii) investigate its performance under various possible scenarios; and (iii) enlarge the suitable set of prior distribution choices for random effects covariance matrices (this provides the data analyst with flexibility in foregoing prior choice in the analysis of multilevel models).

The rest of the article on that account is organized as follows: In Section 2, we briefly mention the well-known datasets containing scalar response and longitudinal covariate measurements and present the data we analyzed in Section 5. Section 3 elaborates how to use G-MVLG as multivariate variance/covariance prior in the setting of interest. Section 4 presents a simulation study comparing the proposed variance prior and the traditional one in joint model setting in terms of relative bias, posterior interquartile range, and standard error estimates. Section 5 uses the proposed approach to determine whether postprandial glucose through the gestation can be used as a prognostic biomarker for obstetric complications in diabetic women. The article ends with a discussion presented in Section 6.

2 Motivating examples

There are many studies in medical research in which the aim lies in determining the association between the outcome observed at a particular endpoint and the trajectory of a covariate measurable over time prior to the endpoint, the most well-known examples being the datasets produced in SWAN (Study of Women's Health Across the Nation, see Sowers et al.[8]) and CARET (Carotene and Retinol Efficacy Trial, primary results of which are given in Omenn et al.,[9] Omenn et al.,[10] and Goodman et al.[11]). Our particular dataset at hand has come from a study conducted in Zekai Tahir Burak Women Health, Care and Research Hospital in Ankara. The aim of the study was to investigate the association between glycated hemoglobin (HbA1c) levels of pregnant women with gestational diabetes mellitus (GDM) recorded over the gestation period and the probability of obstetrics complication. The (log) serum HbA1c of the patients were recorded at the patients antenatal visits. Minimum, median, and maximum number of recordings were two, five, and eight, respectively. The dataset included, for each (i) of 259 women with GDM, (log) serum $HbA1c$ measured at each antenatal visit (i.e. at t_{ij};$HbA1c_{ij}$;continuous), maternal characteristics (maternal age (Age_i;continuous), body mass index of patient before 8 weeks of gestation (BMI_i;continuous), parity (P_i;continuous), GDM history ($GDMH_i$;binary), GDM family history ($GDMFH_i$;binary), hemoglobin concentration > 13 g/dL ($Hb13_i$;binary), macrosomia history (MSH_i;binary)), and the status of obstetric labor complication (OLC_i;binary). For the binary variables, absence is the reference level. The analyses were carried out for GDM patients on insulin and diet treatments separately. In this article, we use the partial dataset consisting of patients on insulin only for illustrative purpose. Complete analyses including different types of obstetrics complications will be published in an obstetric journal.

Figure 1 shows the HbA1c profile of the patients with (right) and without (left) an obstetrics complication. Bold lines in the Figure 1 basically join the average of HbA1c levels of women at time t_{ij}'s. Visual comparison of the two figures offer a preliminary intuition about the association between the HbA1c profile over the gestation and occurrence of obstetrics complication. However, surely it is still difficult to set firmly whether the two figures are statistically significantly (dis)similar. To shed light on whether obstetrics complication is associated with HbA1c profile, controlled for all the other factors, the dataset is analyzed in Section 5 using the Bayesian methods considered in this article.

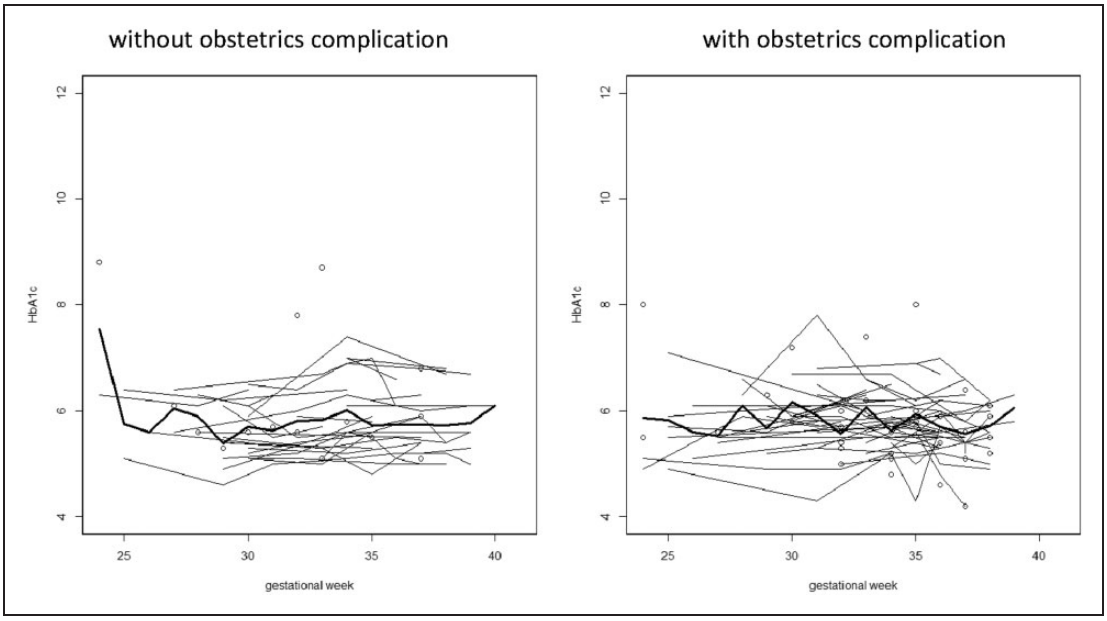

Figure 1. Serum HbA1c versus gestational week for pregnant women with GDM.

3 Joint model and a joint Bayesian analysis

3.1 A Bayesian view of the joint model

We consider the joint model framework of Li et al.[2] for modeling generalized responses on longitudinally measured continuous covariates. We restate the model here from a Bayesian viewpoint. Let Y_i denote the response obtained from subject i, $i = 1, \ldots, n$, at a primary endpoint, W_{ij} be the jth observed longitudinal continuous measurement for subject i taken prior to the primary endpoint of interest, $W_i = \{W_{i1}, W_{i2}, \ldots, W_{in_i}\}$ be the set of these longitudinal measurements with n_i being the number of measurements for subject i, and Z_i be the $1 \times p$ vector of cross-sectional covariates of mixed type measured for subject i including 1 for the intercept. Also let X_i denote the random coefficients ($q \times 1$) characterizing the W_i profile over time. In addition, let β be the $p \times 1$ vector of unknown regression coefficients associated with the cross-sectional covariates and γ be the $q \times 1$ vector of unknown regression coefficients associated with X_i. Below, $U_i = \{U_{i1}, U_{i2}, \ldots, U_{in_i}\}$ be the vector of uncorrelated error terms specified under the conditional independency assumption for W_{ij}s given X_i. Then, for $i = 1, \ldots, n$,

$$\begin{aligned} f(Y_i|Z_i, X_i, \beta, \gamma) &= exp\left(\frac{Y_i\theta_i - b(\theta_i)}{a(\phi)} + c(Y_i, \phi)\right) \\ W_i &= D_i X_i + U_i \end{aligned} \tag{1}$$

$$\begin{aligned} X_i|\mu_X, \Sigma_X &\sim F_X \\ U_i|\sigma_U^2 &\sim F_U \end{aligned} \tag{2}$$

$$\begin{aligned} \beta &\sim F_\beta \\ \gamma &\sim F_\gamma \end{aligned} \tag{3}$$

$$\begin{aligned} \mu_X &\sim F_{\mu_X} \\ \Sigma_X &\sim F_{\Sigma_X} \\ \phi &\sim F_\phi \\ \sigma_U^2 &\sim F_{\sigma_U^2} \end{aligned} \tag{4}$$

where $\theta_i = Z_i^T\beta + X_i^T\gamma$, D_i is the $n_i \times q$ design matrix for time variables, ϕ is the precision parameter, F_X is the distribution assumed for X_is with mean μ_X and unstructured covariance matrix Σ_X, F_U is a suitable distribution with mean 0 and covariance matrix $\sigma_U^2 I_{n_i}$, e.g., $N_{n_i}(0, \sigma_U^2 I_{n_i})$, where N_{n_i} is a multivariate normal distribution of dimension n_i and I_{n_i} is an identity matrix of $n_i \times n_i$. Here, X_i's are latent variables shared by Y_i and W_{ij} models. In the first one, they act as (latent) covariates whereas they are the random coefficients in the second one. As indicated above, the hierarchical structure can be thought of in four distinct levels. First level of the hierarchy, labeled with

(1), concerns with models for the observables (i.e. Y_i and W_i for $i = 1, \ldots, N$) given the unobservables. Second level, labeled with (2), imposes distributional assumptions for the unobservables namely the random effects and the random errors. Level (3) serves either to incorporate any prior knowledge about the corresponding associations (in which case an informative prior emerges) or some initial distributional assumptions about them to start with in the absence of prior knowledge (in which case a noninformative prior emerges). Last level of the hierarchy, level labeled with (4), is where the prior distributions for the parameters of the distributions of the random terms are set up. Focus of our article is the variance and variance–covariance matrix priors in this level.

When the primary-endpoint-outcome, Y_i, is a continuous measurement, which is modeled by, e.g., $Y_i = Z_i^T\beta + X_i^T\gamma + \in_i$, where $\in_i$s are error terms that are independently and identically distributed with mean 0 and variance $\sigma_\in^2$, total marginal variation in Y_i is given by $Var(Y_i) = \gamma^T\Sigma_X\gamma + \sigma_\in^2$. That is, the constituents of the total variation in Y_i are variance and covariance terms in Σ_X as well as $\sigma_\in^2$. Similarly, in the joint model, total marginal covariance in W_i is given by $Var(W_i) = D_i\Sigma_X D_i^T + \sigma_U^2 I_{n_i}$. That is, the constituents of the total covariance in W_i are variance and covariance terms in Σ_X as well as σ_U^2. Impelled by the compositional feature of the total variations in the response variables as such, we propose joint a priori modeling for the error variances and the random effects variance–covariance matrix components.

3.2 A joint Bayesian analysis: Joint prior for error variance and random effects variance–covariance matrix

Commonplace in the Bayesian analysis of random effects models is to consider the (inverse-gamma, inverse-Wishart) pair as priors for error variances and random effects covariance matrix, respectively. The now-known pitfalls of the inverse-gamma prior should discourage the use of it as default variance prior for noninformative Bayesian analysis (see the seminal work of Gelman[12] and more recently Demirhan and Kalaylioglu[6]). Similarly, inverse-Wishart prior, which implies that variance priors along the diagonal have inverse-gamma distributions, may inherit the same drawbacks of inverse-gamma and thus should be used with care in practice. Moreover, inverse-Wishart is restrictive in the sense that it lacks parameters modeling the prior dependencies between the elements of a covariance matrix (Leonard and Hsu[13]). Most recently, Huang and Wand[14] proposed a family of priors for covariance matrices in which direct use of inverse-Wishart is avoided. Alternatively, a covariance or an inverse covariance matrix is decomposed and prior distributions are assigned to the resulting decompositions (e.g. see Chen and Dunson[15] and Cai and Dunson[16]). This is still an active research area and Cholesky, modified Cholesky, spectral decomposition, and variance-correlation decompositions have been studied to formulate covariance matrix priors. Once the entries of the triangular and diagonal matrices resulted from the decomposition are vectorized, univariate and multivariate prior distributions have been considered. Modified Cholesky decomposition is especially useful in dynamic models when covariance matrix entries are to be modeled conditional on covariates (see, e.g. Daniels and Pourahmadi[17]). Our interest lies in priors for unstructured covariance matrices and we consider the basic Cholesky decomposition, in particular, as a convenient choice. Variance–covariance matrix priors based on the basic Cholesky decomposition can be found in Frühwirth-Schnatter and Tüchler,[18] Tüchler,[19] and Congdon.[20] For an extensive account of decomposition-based variance–covariance matrix prior developments, refer to the references listed in Barnard et al.[21] In addition, for a complete history of covariance matrix modeling as well as the advantages of them in particular of Cholesky, refer to Pourahmadi.[22] Our proposed joint prior modeling approach for the error variances and the variance–covariance matrix components in the joint model in Section 3.1. requires Cholesky decomposition of the random effects variance–covariance matrix.

We utilize Cholesky method to decompose the random effects variance–covariance matrix as $\Sigma_{\mathbf{X}} = \mathbf{C}\mathbf{C}^{\mathbf{T}}$, where $\mathbf{C}$ is a lower triangular matrix, vectorize the diagonals and nonzero off-diagonals of C and denote the resulting column vectors by C_1 and C_2, respectively, and eventually consider a joint prior distribution for $(C_1^T, C_2^T, \sigma_\in^2, \sigma_U^2)^T$ if the outcome variable is continuous and for $(C_1^T, C_2^T, \sigma_U^2)^T$ if the outcome variable is either dichotomous or polychotomous. Ultimately, we consider a multivariate distribution for the vector of log-transformed error variances, log-transformed C_1, and untransformed C_2, in particular, G-MVLG, due to its advantages as a robust noninformative prior that are elucidated in Demirhan and Kalaylioglu.[6] Positive definiteness of Σ_X is ensured by imposing positive priors on C_1 whereas those for C_2 are left unconstrained. Assignment of positive priors is accomplished by considering unconstrained priors for $\log(C_1)$. Finally, Σ_X and σ_U^2 priors on level (4) above becomes,

$$(\log(C_1), C_2, \log\phi, \log\sigma_U^2)^T \sim F_{(\delta,\nu,\lambda,\,\boldsymbol{\eta})} \tag{5}$$

where $F_{(\delta,\nu,\boldsymbol{\lambda},\boldsymbol{\eta})}$ represents the G-MVLG distribution indexed by the positive parameter vector $(\delta, \nu, \boldsymbol{\lambda}, \boldsymbol{\eta})$. Once the Gibbs algorithm produces Markov chains each of which is convergent to the corresponding marginal posterior distribution for each parameter in $(\log(C_1), C_2, \log\sigma_U^2)^T$, Markov chains for the components of Σ_X are obtained by the deterministic back transformation,

$$\begin{aligned}\Sigma_{X,ll} &= \sum_{r=1}^{l} c_{lr}^2 \\ \Sigma_{X,kl} &= \sum_{r=1}^{l} c_{kr}c_{lr},\end{aligned} \tag{6}$$

where $\Sigma_{X,ll}$ and $\Sigma_{X,kl}$ are the variances and the covariances, respectively, c_{ll} are the diagonals of C and also the elements of C_1, c_{kl} are the nonzero off-diagonals of C and also the elements of C_2 for $l = 1, \ldots, q,\ k = (l+1), \ldots, q$. Immediate benefits of the approach are (1) use of inverse-gamma and inverse-Wishart are avoided and (2) compositional feature of total variance is accounted for. For each β_i, a slightly flat normal prior is considered. Initial Gibbs runs with noninformative priors for γ_i's resulted in long-lasting autocorrelations, which may be associated with an identifiability issue. To avoid identifiability problem related with γ coefficients, weakly informative priors are considered. For the variance parameters, independent priors (inverse-gamma or uniform, inverse-Wishart) and joint prior approach (based on G-MVLG) are considered as described earlier for comparative purposes. For each μ_X, diffuse prior is used. For the fixed hyperparameters of all these prior distributions, see Section 3. Joint posterior distribution of model parameters and the latent variables for generalized Y is,

$$\begin{aligned}f(\beta, \gamma, \phi, \sigma_U^2, \Sigma_X, \mu_X, X|Y, W, Z, D) \propto \quad & L(\beta, \gamma, \phi, \sigma_U^2, \Sigma_X, \mu_X|X, Y, W, Z, D) \\ & \times f_\beta(\beta) f_\gamma(\gamma) f_{\phi,\sigma_U^2,\Sigma_X}(\phi, \sigma_U^2, \Sigma_X)\end{aligned},$$

where $f_\beta(\beta), f_\gamma(\gamma)$ and $f_{\phi,\sigma_U^2,\Sigma_X}(\phi, \sigma_U^2, \Sigma_X)$ are priors and

$$\begin{aligned}L(\beta, \gamma, \phi, \sigma_U^2, \Sigma_X, \mu_X|X, Y, W, Z, D) = \quad & f(Y|X, W, Z, D, \beta, \gamma, \phi) \\ & \times f(W|X, D, \sigma_U^2) f(X|\mu_X, \Sigma_X)\end{aligned}$$

is the complete data likelihood components of which are readily written out based on the model described in Section 3. Gibbs sampling via OpenBUGS is used to obtain the marginal posterior distributions of the parameters.

4 Simulation study

In this simulation study, analysis with the proposed joint variance prior approach is examined and compared with the standard approach in terms of bias in and efficiency of posterior estimators. Aim of the simulation study is to investigate the following questions, in particular. Does joint variance approach improve posterior inference on regression coefficients (β and γ) and variance components relative to the standard approach? How do they compare with respect to the sample size and the true association between the cross-sectional outcome and profile of the longitudinal covariate? Simulation experiment is controlled for sample size (n) and the strength of the aforementioned association (γ_1 and γ_2). Different scenarios encompassing moderate to large sample sizes ($n = 250$ and $n = 500$) and positive to negative associations are studied. Our data producing mechanism resembles a natural setting and is as follows. Random effects $X_i = (X_{i1}, X_{i2})$ for each $i = 1, \ldots, n$ are generated from $N_2(\mu_X, \Sigma_X)$, where $\mu_X = (0.5, 0.5)^T$ and $\Sigma_X = \begin{pmatrix} 1 & -0.2 \\ -0.2 & 0.64 \end{pmatrix}$. Given the random effects, each binary response Y_i is generated from a Bernoulli distribution with success probability p_i, where $logit(p_i) = \beta + \gamma_1 X_{i1} + \gamma_2 X_{i2}$. For (γ_1, γ_2), three different scenarios are considered implying negative association $(-3,-2)$, positive association (3,2), and near-zero association (0.3,0.2) between the odds of $Y_i = 1$ and the longitudinal profile of W_i. Two prevalence scenarios are considered for $P(Y = 1)$ illustrating a rare medical problem ($P(Y = 1) = 0.1$) and a rather common problem ($P(Y = 1) = 0.5$). True β corresponding to resulting six scenarios are given in Tables 1–6. Longitudinal covariates are generated from $W_{ij} = X_{i1} + X_{i2}t_{ij} + U_{ij}$ for $i = 1, \ldots, n,\ j = 1, \ldots, n_i$, where t_{ij} is the time point at which W_{ij} is measured and $U_{ij}^{iid} \sim N(0, \sigma_U^2)$ (iid = independently and identically distributed). The number of longitudinal measurements, n_is, and

Table 1. Summary statistics for posterior medians.

		Relative bias		Interquartile range		MC standard error	
n	True parameter	G-MVLG	(IG,IW)	G-MVLG	(IG,IW)	G-MVLG	(IG,IW)
250	$\beta = -4.6$	0.129	0.143	1.724	1.752	1.314	1.317
	$\gamma_1 = 3$	0.167	0.194	1.318	1.352	0.931	0.943
	$\gamma_2 = 2$	0.106	0.119	0.845	0.855	0.708	0.705
	$\sigma_U^2 = 0.5$	0.005	0.011	0.035	0.036	0.027	0.027
	$\text{Var}(X_{i0}) = 1$	0.017	−0.026	0.191	0.223	0.134	0.132
	$\text{Var}(X_{i1}) = 0.64$	0.019	−0.001	0.079	0.077	0.055	0.054
	$\text{Cov}(X_{i0}, X_{i1}) = -0.2$	0.057	0.042	0.089	0.161	0.069	0.068
500	$\beta = -4.6$	0.067	0.075	1.209	1.233	0.783	0.783
	$\gamma_1 = 3$	0.080	0.094	0.928	0.947	0.641	0.644
	$\gamma_2 = 2$	0.069	0.078	0.599	0.609	0.393	0.394
	$\sigma_U^2 = 0.5$	0.005	0.007	0.026	0.026	0.017	0.017
	$\text{Var}(X_{i0}) = 1$	0.0002	−0.021	0.134	0.153	0.101	0.100
	$\text{Var}(X_{i1}) = 0.64$	0.024	0.014	0.056	0.055	0.042	0.041
	$\text{Cov}(X_{i0}, X_{i1}) = -0.2$	−0.019	−0.025	0.063	0.109	0.042	0.042

Note: True P $(Y = 1|X) = 0.1$.

Table 2. Summary statistics for posterior medians.

		Relative bias		Interquartile range		MC standard error	
n	True parameter	G-MVLG	(IG,IW)	G-MVLG	(IG,IW)	G-MVLG	(IG,IW)
250	$\beta = 0.3$	0.123	0.251	0.546	0.561	0.356	0.362
	$\gamma_1 = -3$	0.116	0.142	1.302	1.336	0.842	0.849
	$\gamma_2 = -2$	0.080	0.096	0.839	0.854	0.580	0.580
	$\sigma_U^2 = 0.5$	0.007	0.013	0.036	0.036	0.026	0.026
	$\text{Var}(X_{i0}) = 1$	0.012	−0.030	0.190	0.226	0.130	0.127
	$\text{Var}(X_{i1}) = 0.64$	0.018	−0.001	0.079	0.076	0.054	0.053
	$\text{Cov}(X_{i0}, X_{i1}) = -0.2$	0.054	0.039	0.089	0.162	0.068	0.067
500	$\beta = 0.3$	0.074	0.137	0.387	0.395	0.272	0.276
	$\gamma_1 = -3$	0.091	0.105	0.969	0.987	0.703	0.715
	$\gamma_2 = -2$	0.049	0.058	0.613	0.621	0.431	0.436
	$\sigma_U^2 = 0.5$	0.005	0.008	0.026	0.026	0.017	0.017
	$\text{Var}(X_{i0}) = 1$	−0.001	−0.022	0.134	0.153	0.103	0.102
	$\text{Var}(X_{i1}) = 0.64$	0.024	0.014	0.056	0.055	0.042	0.042
	$\text{Cov}(X_{i0}, X_{i1}) = -0.2$	−0.015	−0.022	0.063	0.109	0.043	0.042

Note: True P $(Y = 1|X) = 0.1$.

measurement times, t_{ij}s are obtained as follows: Mimicking the GDM study, a study period of 36 weeks is considered and $n_i = 5$ time points are generated randomly from a discrete uniform distribution (1, 36) for all i (balanced). Then to produce imbalanced longitudinal data, each t_{ij} is made arbitrarily missing with probability 0.05. Simulations are repeated 100 times.

Priors are $\beta \sim N(0, 100)$, $(\gamma_1, \gamma_2)^T \sim N_2((0,0)^T, 10I_2)$, and $\mu_X \sim N_2((0,0)^T, 10I_2)$. Each simulated dataset is analyzed using both of the approaches on variance parameters. In the first approach, $1/\sigma_U^2 \sim Ga(2.02, 1.49)$ and $\Sigma_X^{-1} \sim Wishart(R, 5.06)$, where $R = \begin{pmatrix} 0.88 & -0.3 \\ -0.3 & 0.67 \end{pmatrix}$. In the second approach, $(\log(C_1), C_2, \log\sigma_U^2)^T \sim G - MVLG(\delta, \nu, \lambda, \eta)$, where $\delta = 0.3$, $\nu = 1.42$, $\lambda = (0.3, 0.3, 0.3, 0.4)^T$, and $\eta = (0.25, 0.35, 0.25, 0.1)^T$. These hyperparameter values are chosen so that the two approaches impose similar degree of prior uncertainty on the variance parameters. For each approach, our OpenBUGS codes are run to

Table 3. Summary statistics for posterior medians.

		Relative bias		Interquartile range		MC standard error	
n	True parameter	G-MVLG	(IG,IW)	G-MVLG	(IG,IW)	G-MVLG	(IG,IW)
250	$\beta = -2.5$	0.045	0.047	0.523	0.529	0.394	0.396
	$\gamma_1 = 0.3$	0.060	0.083	0.420	0.428	0.314	0.321
	$\gamma_2 = 0.2$	−0.121	−0.107	0.397	0.399	0.325	0.326
	$\sigma_U^2 = 0.5$	0.006	0.013	0.037	0.037	0.027	0.027
	$\text{Var}(X_{i0}) = 1$	0.009	−0.035	0.194	0.235	0.139	0.136
	$\text{Var}(X_{i1}) = 0.64$	0.019	−0.000	0.079	0.077	0.055	0.053
	$\text{Cov}(X_{i0}, X_{i1}) = -0.2$	0.070	0.056	0.090	0.167	0.070	0.068
500	$\beta = -2.5$	0.019	0.020	0.351	0.353	0.260	0.262
	$\gamma_1 = 0.3$	−0.036	−0.025	0.282	0.285	0.211	0.215
	$\gamma_2 = 0.2$	0.024	0.033	0.273	0.274	0.186	0.187
	$\sigma_U^2 = 0.5$	0.000	0.003	0.025	0.026	0.020	0.020
	$\text{Var}(X_{i0}) = 1$	−0.000	−0.022	0.136	0.155	0.097	0.096
	$\text{Var}(X_{i1}) = 0.64$	0.003	−0.007	0.055	0.054	0.042	0.042
	$\text{Cov}(X_{i0}, X_{i1}) = -0.2$	−0.005	−0.012	0.063	0.112	0.047	0.046

Note: True P $(Y = 1|X) = 0.1$.

Table 4. Summary statistics for posterior medians.

		Relative bias		Interquartile range		MC standard error	
n	True parameter	G-MVLG	(IG,IW)	G-MVLG	(IG,IW)	G-MVLG	(IG,IW)
250	$\beta = -2.5$	0.148	0.169	1.085	1.109	0.688	0.685
	$\gamma_1 = 3$	0.186	0.213	1.332	1.370	0.839	0.839
	$\gamma_2 = 2$	0.116	0.131	0.827	0.843	0.501	0.497
	$\sigma_U^2 = 0.5$	0.011	0.017	0.036	0.036	0.027	0.027
	$\text{Var}(X_{i0}) = 1$	−0.033	−0.075	0.186	0.239	0.138	0.135
	$\text{Var}(X_{i1}) = 0.64$	0.014	−0.005	0.079	0.077	0.052	0.051
	$\text{Cov}(X_{i0}, X_{i1}) = -0.2$	−0.028	−0.040	0.088	0.166	0.063	0.062
500	$\beta = -2.5$	0.126	0.138	0.801	0.815	0.653	0.655
	$\gamma_1 = 3$	0.143	0.157	0.967	0.985	0.789	0.791
	$\gamma_2 = 2$	0.107	0.116	0.616	0.623	0.463	0.469
	$\sigma_U^2 = 0.5$	0.005	0.008	0.025	0.026	0.020	0.020
	$\text{Var}(X_{i0}) = 1$	−0.007	−0.028	0.134	0.156	0.104	0.103
	$\text{Var}(X_{i1}) = 0.64$	0.009	−0.002	0.055	0.055	0.041	0.040
	$\text{Cov}(X_{i0}, X_{i1}) = -0.2$	0.011	0.003	0.062	0.112	0.040	0.040

Note: True P $(Y = 1|X) = 0.5$.

obtain two Markov chains. First 5000 iterations of each chain are discarded and every 10th of next 25,000 iterations are used for posterior calculations (thinning, chain size, and convergence were determined based on the autocorrelation plots, Monte Carlo (MC) standard errors, and estimated potential scale reduction factors (Gelman and Rubin[23])).

Posterior medians are used to estimate the parameters. We use posterior interquartile range as a measure of efficiency. G-MVLG and (inverse-gamma, inverse-Wishart)-based approaches are compared based on relative bias and posterior interquartile range. MC estimates of these quantities along with MC standard error estimates of posterior medians are given in Tables 1–6. In the tables, (inverse-gamma, inverse-Wishart) is denoted by (IG,IW). From the tables, we see that G-MVLG joint prior consistently leads to smaller relative bias and narrower interquartile range in response model coefficient estimators, the parameters of the main inferential concern. G-MVLG and (IG,IW) have comparable MC standard error estimates which makes the comparison of relative biases and interquartile ranges meaningful.

Table 5. Summary statistics for posterior medians.

		Relative bias		Interquartile range		MC standard error	
n	True parameter	G-MVLG	(IG,IW)	G-MVLG	(IG,IW)	G-MVLG	(IG,IW)
250	$\beta = 2.5$	0.198	0.221	1.125	1.150	0.709	0.714
	$\gamma_1 = -3$	0.214	0.241	1.351	1.390	0.828	0.834
	$\gamma_2 = -2$	0.190	0.205	0.872	0.886	0.570	0.569
	$\sigma_U^2 = 0.5$	0.008	0.013	0.035	0.036	0.029	0.029
	$\text{Var}(X_{i0}) = 1$	0.006	−0.035	0.189	0.222	0.123	0.119
	$\text{Var}(X_{i1}) = 0.64$	0.022	0.002	0.079	0.077	0.059	0.061
	$\text{Cov}(X_{i0}, X_{i1}) = -0.2$	−0.008	−0.020	0.088	0.158	0.058	0.060
500	$\beta = 2.5$	0.075	0.088	0.757	0.772	0.508	0.511
	$\gamma_1 = -3$	0.099	0.114	0.918	0.940	0.634	0.634
	$\gamma_2 = -2$	0.069	0.079	0.585	0.595	0.414	0.418
	$\sigma_U^2 = 0.5$	0.005	0.008	0.026	0.026	0.017	0.017
	$\text{Var}(X_{i0}) = 1$	0.0001	−0.022	0.135	0.153	0.101	0.099
	$\text{Var}(X_{i1}) = 0.64$	0.024	0.014	0.056	0.055	0.042	0.042
	$\text{Cov}(X_{i0}, X_{i1}) = -0.2$	−0.014	−0.021	0.063	0.109	0.042	0.042

Note: True P $(Y = 1|X) = 0.5$.

Table 6. Summary statistics for posterior medians.

		Relative bias		Interquartile range		MC standard error	
n	True parameter	G-MVLG	(IG,IW)	G-MVLG	(IG,IW)	G-MVLG	(IG,IW)
250	$\beta = -0.25$	−0.059	−0.043	0.267	0.268	0.192	0.194
	$\gamma_1 = 0.3$	0.108	0.128	0.253	0.260	0.207	0.212
	$\gamma_2 = 0.2$	−0.074	−0.061	0.236	0.236	0.183	0.183
	$\sigma_U^2 = 0.5$	−0.001	0.005	0.036	0.036	0.031	0.031
	$\text{Var}(X_{i0}) = 1$	−0.015	−0.059	0.190	0.237	0.126	0.123
	$\text{Var}(X_{i1}) = 0.64$	0.020	−0.0003	0.079	0.077	0.061	0.059
	$\text{Cov}(X_{i0}, X_{i1}) = -0.2$	−0.026	−0.039	0.089	0.165	0.063	0.061
500	$\beta = -0.25$	−0.008	0.0001	0.183	0.184	0.125	0.126
	$\gamma_1 = 0.3$	−0.058	−0.048	0.171	0.172	0.124	0.126
	$\gamma_2 = 0.2$	0.051	0.056	0.162	0.163	0.114	0.115
	$\sigma_U^2 = 0.5$	0.004	0.007	0.026	0.026	0.017	0.017
	$\text{Var}(X_{i0}) = 1$	−0.002	−0.024	0.135	0.156	0.103	0.102
	$\text{Var}(X_{i1}) = 0.64$	0.024	0.014	0.056	0.055	0.042	0.041
	$\text{Cov}(X_{i0}, X_{i1}) = -0.2$	−0.014	−0.020	0.063	0.110	0.042	0.042

Note: True P $(Y = 1|X) = 0.5$.

5 Association between HbA1c and obstetric labor complication among women with diabetes mellitus

In this section, GDM dataset described in Section 2 is revisited. We model the data using the joint model and employ both standard and joint Bayesian approach for its analysis.

Number of longitudinal serum $HbA1c$ measurements (n_i) varying between only three and eight per subject, unaligned, and not having a particular functional form over the weeks (as seen in Figure 1) motivate a simple regression for the longitudinal model in (1) only with random intercept and slope terms. A generalized linear regression is used for the probability of OLC, in which longitudinal log serum HbA1c are accounted for by

including covariates that are the random coefficients from the longitudinal model. Hence, the mixed effects response models considered are,

$$\begin{aligned} probit(P(OLC_i = 1|\text{covariates}_i)) &= \beta_1 + \beta_2 Age_i + \beta_3 BMI_i + \beta_4 P_i + \beta_5 GDMH_i \\ &\quad + \beta_6 GDMFH_i + \beta_7 Hb13_i + \beta_8 MSH_i + \gamma_1 X_{i1} + \gamma_2 X_{i2} \\ log(HbA1c_{ij}) &= X_{i1} + X_{i2}\left(\frac{t_{ij}-\bar{t}}{S_t}\right) + U_{ij} \end{aligned}$$

where $\bar{t}$ and S_t are the mean and the standard deviation of all t_{ij}'s, $i = 1,\ldots,259$, $j = 1,\ldots,n_i$. Here, γ_1 and γ_2 together represent the degree of association between the probability of OLC and the mother's log serum $HbA1c$ profile over the gestation period. In particular, γ_2 represents the association between the probability of OLC and the rate of change in log serum $HbA1c$ in time and that being equal to 0 indicates OLC status is not related with the slope of mother's log serum $HbA1c$ across the gestation. γ_1, on the other hand, represents the association between the probability of OLC and the mother's mean log serum $HbA1c$. Of the two, attention centers about the inference on γ_2 as it tells whether the way the mother's log serum $HbA1c$ changes across the gestational age can be used as a biomarker for OLC.

Probit link is preferred for computational benefits following Albert and Chib.[24] Since X_{i1} and X_{i2} are latent, one has to consider a population distribution for them, so that Gibbs sampling can integrate them out. For the data in which all the subjects exhibit a linear trend in a longitudinal plot, one can establish such a distribution by fitting individual regressions for each subject, collecting all the individual intercept and slope estimators together (i.e. estimators of X_{i1}'s and X_{i2}'s for $i = 1,\ldots,259$) and fit a distribution to those. For the data in which longitudinal profiles vary considerably across individuals, as in the dataset analyzed here, a reasonable population model is assumed instead such as a bivariate normal ($N_2(\mu_X, \Sigma_X)$) or a bivariate t ($t_2(\mu_X, \Sigma_X, 4)$).

Priors are $\beta_1 \sim N(0, 100)$, $\beta_p \sim N(0, 10^3)$ for $p = 2, \ldots, 8$, $\gamma_q \sim N(0, 10)$ for $q = 1$, 2, and $\mu_X \sim N_2((0,0)^T, 100I_2)$. β_1 is assigned a slightly more informative prior as our initial Gibbs runs indicated a large autocorrelation on earlier lags, which may be due to an identifiability issue in the sample. For the usual independent priors for the error variance and the random effects covariance matrix, $1/\sigma_U^2 \sim Ga(0.01, 0.01)$ and $\Sigma_X^{-1} \sim Wishart(0.001I_2, 2)$. For joint variance prior $(\log(C_1), C_2, \log\sigma_U^2)^T \sim G - MVLG(0.7, 1.42, \lambda, \eta)$ with $\lambda = (0.3, 0.3, 0.3, 0.4)^T$, $\eta = (0.25, 0.35, 0.25, 0.1)^T$. Both of these priors are slightly noninformative and parameters are selected, so that they have similar uncertainty on the variance components. Convergence is ensured and the pseudo-convergence is avoided by running one long chain (as advised by Brooks et al.[25]). First, 75,000 iterations are burnt and every 50th of next 925,000 iterations resulting in 18,500 iterations in total are hired for the posterior inference. A sample OpenBUGS code for G-MVLG is provided below.

Posterior means and the 95% quantile-based posterior intervals for the probit regression coefficients are given in Table 7. First column lists the covariates. X_1 and X_2 are the mean $HbA1c$ (intercept of the $HbA1c$ process) and change in $HbA1c$ (slope of the $HbA1c$ process), respectively. Accordingly, average $HbA1c$ level during the gestation period is positively associated with probability of OLC (neither of the related 95% posterior intervals encompass zero), adjusted for the other factors. Also, we see that posterior inference obtained from Bayesian

Table 7. Posterior means and the 95% posterior intervals of the coefficients.

	Joint variance prior (G-MVLG)		Independent variance priors (IG,IW)	
Covariates	$X_i^T \sim N_2(\mu_X, \Sigma_X)$	$X_i^T \sim t_2(\mu_X, \Sigma_X, 4)$	$X_i^T \sim N_2(\mu_X, \Sigma_X)$	$X_i^T \sim t_2(\mu_X, \Sigma_X, 4)$
Intercept	−7.81 (−13.50, −2.51)	−7.69 (−13.13, −2.44)	−9.93 (−17.28, −2.84)	−9.84 (−17.83, −2.79)
Age	0.002 (−0.06, 0.06)	0.003 (−0.06, 0.06)	−0.003 (−0.07, 0.06)	0.0007 (−0.07, 0.07)
BMI	0.003 (−0.06, 0.06)	0.004 (−0.06, 0.07)	0.006 (−0.07, 0.08)	0.006 (−0.07, 0.08)
Parity	0.23 (−0.02, 0.50)	0.23 (−0.02, 0.49)	0.26 (−0.02, 0.55)	0.26 (−0.03, 0.56)
GDMH	0.43 (−0.20, 1.05)	0.43 (−0.20, 1.06)	0.54 (−0.14, 1.24)	0.55 (−0.16, 1.26)
GDMFH	0.08 (−0.66, 0.78)	0.08 (−0.64, 0.79)	0.05 (−0.76, 0.82)	0.05 (−0.78, 0.84)
Hb13	0.52 (−0.09, 1.15)	0.52 (−0.11, 1.14)	0.56 (−0.11, 1.28)	0.58 (−0.10, 1.31)
MSH	−1.35 (−2.66, −0.26)	−1.35 (−2.62, −0.25)	−1.46 (−2.89, −0.28)	−1.48 (−2.89, −0.28)
X_1	3.44 (0.73, 6.39)	3.34 (0.68, 6.14)	4.60 (0.77, 8.50)	4.47 (0.65, 8.72)
X_2	0.37 (−5.83, 6.50)	0.68 (−5.65, 6.79)	0.85 (−5.37, 7.14)	0.92 (−5.39, 7.52)

analysis with joint variance priors and that with independent variance priors are similar with one notable difference: length of the posterior intervals obtained from the former approach is invariably shorter than the latter.

Conditional independence of W_{ij}'s is tested through testing independence of U_{ij}'s. Wald–Wolfowitz runs test of randomness is used and p values are 0.56 and 0.91 for models with $X_i^T \sim N_2(\mu_X, \Sigma_X)$ and $X_i^T \sim t_2(\mu_X, \Sigma_X, 4)$, respectively, indicating no evidence against the conditional independence assumption.

```
#Prior modeling for
#(log(C1),C2,log(sigma2u)) ~ G-MVLG(delta,nu,lambdas,mus).
#Below (theta[1],theta[2],theta[3],theta[4]) =
# (log(C1),C2,log(sigma2u))

auu <- 1/sigma2u
sigma2u <- exp(theta[4])
invcovx[1:2,1:2] <- inverse(covx[1:2,1:2])
covx[1,1] <- sigma2.11
covx[1,2] <- sigma2.12
covx[2,1] <- sigma2.12
covx[2,2] <- sigma2.22

#for positive definitness of covx
sigma2.11 <- l11 * l11
sigma2.12 <- l21 * l11
sigma2.22 <- l21*l21 + l22*l22

l11 <- exp(theta[1])
l21 <- theta[2]
l22 <- exp(theta[3])

theta[1] ~ dflat()
theta[2] ~ dflat()
theta[3] ~ dflat()
theta[4] ~ dflat()

dummy <- 0
dummy ~ dloglik(phi)

#G-MVLG log(pdf); D is a finite upper bound approximating
#the infinite series in the pdf.
phi <- nu * log(delta) + log(sum(v[1:D]))
for(i in 1: D) {
v[i] <- (v1[i] / v2[i]) * v3[i]
v1[i] <- pow((1-delta),nu) * (mus[1]
* pow(lambdas[1],-nu-i)) * (mus[2]
* pow(lambdas[2],-nu-i)) * (mus[3]
* pow(lambdas[3],-nu-i)) * (mus[4]
* pow(lambdas[4],-nu-i))
v2[i] <- (pow(exp(loggam(nu+i)),3)
* exp(loggam(nu)) * exp(logfact(i)))
v3[i] <- exp((nu+i) * (mus[1]*theta[1]
+ mus[2]*theta[2] + mus[3]*theta[3]
+ mus[4]*theta[4]) - ((1/lambdas[1])
* exp(mus[1]*theta[1]) + (1/lambdas[2])
* exp(mus[2]*theta[2]) + (1/lambdas[3])
* exp(mus[3]*theta[3]) + (1/lambdas[4])
* exp(mus[4]*theta[4])))}
```

6 Conclusion

In this article, we have proposed modeling the variance components of a random effects model a priori jointly. Simulation experiment revealed that proposed approach has better posterior qualities compared with standard Bayesian analysis. This is also reflected in the application section. Our approach provides a nonconjugate alternative for prior modeling of variance parameters in random effects models. G-MVLG probability density function is a nonstandard one, requires an infinite series (see Demirhan and Hamurkaroglu[7]), and not readily available in OpenBUGS. However, it can be coded using zeros trick as seen in the application section. Computational complexities may arise if the dimension of the vector of variance components to be modeled jointly is large, which may be the case, e.g., when W_{ij} model is a polynomial of a large order. Given the advantages (reduced bias, improved efficiency, and easy coding), it, nevertheless, seems to be a suitable prior modeling for variance and covariance components especially when the dimension of random/latent terms is moderate. Extension of the joint Bayesian approach pursued herein to the analysis of response measured at a primary endpoint and multiple longitudinal covariates of mixed type should be straightforward. The idea presented here can easily be extended to a wider class of random effects models.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Wang CY, Wang N and Wang S. Regression analysis when covariates are regression parameters of a random effects model for observed longitudinal measurements. *Biometrics* 2000; **56**: 487–495.
2. Li E, Zhang D and Davidian M. Conditional estimation for generalized linear models when covariates are subject-specific parameters in a mixed model for longitudinal measurements. *Biometrics* 2004; **60**: 1–7.
3. Li E, Wang N and Wang NY. Joint models for a primary endpoint and multiple longitudinal covariate processes. *Biometrics* 2007; **63**: 1068–1078.
4. Li E, Zhang D and Davidian M. Likelihood and pseudo-likelihood methods for semiparametric joint models for a primary endpoint and longitudinal data. *Comput Stat Data Anal* 2007; **51**: 5776–5790.
5. Horrocks J and van den Heuvel MJ. Prediction of pregnancy: A joint model for longitudinal and binary data. *Bayesian Anal* 2009; **4**: 523–538.
6. Demirhan H and Kalaylioglu Z. Joint prior distributions for variance parameters in Bayesian analysis of normal hierarchical models. *J Multivariate Anal* 2014; **15**: 163–174.
7. Demirhan H and Hamurkaroglu C. On a multivariate log-gamma distribution and the use of the distribution in Bayesian analysis. *J Stat Plan Inference* 2011; **141**: 1141–1152.
8. Sowers MR, Finkelstein J, Ettinger B, et al. The association of endogenous hormone concentrations and bone mineral density measures in pre- and peri-menopausal women of four ethnic groups: SWAN. *Osteoporosis Int* 2003; **14**: 44–52.
9. Omenn GS, Goodman GE, Thornquist MD, et al. Effects of a combination of beta-carotene and vitamin A on lung cancer and cardiovascular disease. *NEJM* 1996; **334**: 1150–1155.
10. Omenn GS, Goodman GE, Thornquist MD, et al. Risk factors for lung cancer and for intervention effects in CARET, the beta-carotene and Retinol Efficacy Trial. *JNCI* 1996; **88**: 1550–1559.
11. Goodman GE, Thornquist MD, Balmes J, et al. The beta-carotene and retinol efficacy trial: Incidence of lung cancer and cardiovascular disease mortality during 6-year follow-up after stopping β-carotene and retinol supplements. *JNCI* 2004; **96**: 1743–1750.
12. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Anal* 2006; **1**: 515–553.
13. Leonard T and Hsu JSJ. Bayesian inference for a covariance matrix. *Ann Stat* 1992; **20**: 1669–1696.
14. Huang A and Wand MP. Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Anal* 2013; **8**: 439–452.
15. Chen Z and Dunson DB. Random effects selection in linear mixed models. *Biometrics* 2003; **59**: 762–769.
16. Cai B and Dunson DB. Bayesian covariance selection in generalized linear mixed models. *Biometrics* 2006; **62**: 446–457.
17. Daniels MJ and Pourahmadi M. Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika* 2002; **89**: 553–566.
18. Frühwirth-Schnatter S and Tüchler R. Bayesian parsimonious covariance estimation for hierarchical linear mixed models. *Stat Comput* 2008; **18**: 1–13.

19. Tüchler R. Bayesian variable selection for logistic models using auxiliary mixture sampling. *J Comp Graph Stat* 2008; **7**: 76–94.
20. Congdon P. *Applied Bayesian hierarchical methods*. Boca Raton, FL: Chapman and Hall/CRC, 2010.
21. Barnard J, McCulloch R and Meng XL. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Stat Sin* 2000; **10**: 1281–1311.
22. Pourahmadi M. Covariance estimation: The GLM and regularization perspectives. *Stat Sci* 2011; **3**: 369–387.
23. Gelman A and Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci* 1992; **7**: 457–511.
24. Albert JH and Chib S. Bayesian analysis of binary and polychotomous response data. *JASA, Theory Methods* 1993; **88**: 669–679.
25. Brooks S, Gelman A, Jones GL, et al. *Handbook of Markov chain Monte Carlo*. Boca Raton, FL: Chapman and Hall CRC, 2011.