

Bayesian modeling of random effects precision/covariance matrix in cumulative logit random effects models

Jiyeong Kim^a, Insuk Sohn^b, Keunbaik Lee^{1,a}

^aDepartment of Statistics, Sungkyunkwan University, Korea;

^bCenter for Biostatistics and Clinical Epidemiology, Samsung Medical Center, Korea

Abstract

Cumulative logit random effects models are typically used to analyze longitudinal ordinal data. The random effects covariance matrix is used in the models to demonstrate both subject-specific and time variations. The covariance matrix may also be homogeneous; however, the structure of the covariance matrix is assumed to be homoscedastic and restricted because the matrix is high-dimensional and should be positive definite. To satisfy these restrictions two Cholesky decomposition methods were proposed in linear (mixed) models for the random effects precision matrix and the random effects covariance matrix, respectively: modified Cholesky and moving average Cholesky decompositions. In this paper, we use these two methods to model the random effects precision matrix and the random effects covariance matrix in cumulative logit random effects models for longitudinal ordinal data. The methods are illustrated by a lung cancer data set.

Keywords: generalized linear mixed models, modified Cholesky decomposition, heterogeneity, positive definiteness, autoregressive, moving-average

1. Introduction

Longitudinal data are measured from each subject repeatedly over time (Diggle *et al.*, 2002). Therefore, outcomes from the same subject are correlated and the correlation is taken into account for the proper estimation of the coefficients of covariates. In the analysis of the longitudinal categorical data, we typically use generalized linear mixed models (GLMMs) (Breslow and Clayton, 1993). The maximum likelihood estimators of the model parameters in the GLMMs are inconsistent when the random effects distribution is misspecified (Litiere *et al.*, 2007, 2008). However the random effects distribution provides random intercepts and slopes for random variation of subjects without the serial correlation of repeated outcomes such as the autoregressive (AR) structure. In this paper, we consider multivariate normal distribution for the serial dependence of repeated outcomes.

In the multivariate normal distribution for the random effects, random effects covariance matrix explains both subject-specific and time variations. Even though the covariance matrix is heterogeneous, one usually assumes the simple structure of the covariance matrix such as a homogeneous AR(1) covariance matrix to satisfy its positive definiteness. However, it is a strong assumption and the estimated values of the fixed effects in GLMMs under the assumption of homogeneous covariance matrix can be biased (Heagerty and Kurland, 2001).

Recently the (random effects) covariance matrix is assumed to be heterogeneous depending on the covariates and to be a high-dimensional structure. To satisfy these assumptions, two Cholesky

¹ Corresponding author: Department of Statistics, Sungkyunkwan University, 25-2, Sungkyunkwan-ro, Jongno-gu, Seoul 03063, Korea. E-mail: keunbaik@skku.edu

decompositions of the matrix were proposed: modified Cholesky decomposition (MCD) and moving average Cholesky decomposition (MACD). MCD decomposes the inverse of covariance matrix (*precision matrix*) for linear models into generalized autoregressive parameters (GARPs) and innovation variances (IVs) (Pourahmadi, 1999, 2000). The GARPs are dependence parameters that explain the serially dependence from the previous outcomes and the GARPs can be set on the AR structure of the covariance matrix. The IVs are variance parameters and positive IVs guarantee the positive definiteness of the covariance matrix. Bayesian modeling using MCD were proposed in linear mixed models (Daniels and Pourahmadi, 2002; Daniels and Zhao, 2003). Pan and MacKenzie (2003, 2006) used the decomposition to deal with unbalanced longitudinal data and to address joint mean-covariance estimation for linear mixed models. Lee *et al.* (2012) used MCD for the random effects precision matrix in the GLMMs and Lee (2013) proposed Bayesian modeling for the random effects precision matrix of the GLMM to accommodate longitudinal binary data using MCD.

In MACD the *covariance matrix* for linear models is decomposed into generalized moving average parameters (GMAPs) and IVs instead of the precision matrix (Zhang and Leng, 2012). Similar to MCD, the positive IVs guarantee the positive definiteness of the covariance matrix. Lee and Yoo (2014) proposed a Bayesian random effects logistic regression model using MACD for the random effects covariance matrix to analyze longitudinal binary data. Recently, Kim and Lee (2015) surveyed several approaches for modeling of the random effects covariance matrix including MCD and MACD.

This paper analyzes longitudinal ordinal data. We will now briefly describe several models for longitudinal ordinal data. McCullagh (1980) proposed the cumulative logit model for independent ordinal data. Liu and Agresti (2005) presented a general overview of models for (clustered) ordinal categorical data. Most models for correlated ordinal data used random effects to explain the correlation of responses. Gibbons and Hedeker (1997) proposed both the probit and logistic models with random-effects for longitudinal ordinal data. Lee and Daniels (2008) proposed likelihood-based marginalized models for longitudinal ordinal data and the marginalized models were extended to accommodate bivariate longitudinal ordinal data (Lee *et al.*, 2013). Recently, Nooraee *et al.* (2014) reviewed the generalized estimating equation (GEE) method with several software packages (R, SAS, SPSS) for longitudinal ordinal data to provide a population-averaged effects of covariates.

In this paper, we propose cumulative logit models with heterogeneous random effects covariance matrix to analyze longitudinal ordinal data. To model the random effects precision matrix and the random effects covariance matrix in the cumulative logit model, we respectively use MCD and MACD. We also use Bayesian methodology for the estimation of parameters. This paper is organized as follows. In Section 2, we present our motivating dataset. In Section 3, we present Bayesian cumulative logit random effects model for longitudinal ordinal data and we propose models for the random effects precision/covariance matrix. In Section 4, we examine the bias of estimates on parameters to the misspecification of the precision/covariance matrix. In Section 5, we analyze real data from a longitudinal study on lung cancer. Finally, we provide conclusions in Section 6.

2. Lung cancer study

A longitudinal study on lung cancer was designed to compare the response rate (RR) (the percentage of patients whose cancer shrinks or disappears after treatment) (Kim *et al.*, 2012). The study was designed as a prospective randomized non-comparative parallel longitudinal study in a single institute. A total of 96 (48 per arm) patients with lung cancer were randomly assigned to two arms (gefitinib and erlotinib). The main goal of the study was to evaluate the RR for each arm. The RRs were 47.9% (95% CI, 33.8%–62.0%) in gefitinib and 39.6% (95% CI, 25.7%–53.4%) in erlotinib. Median

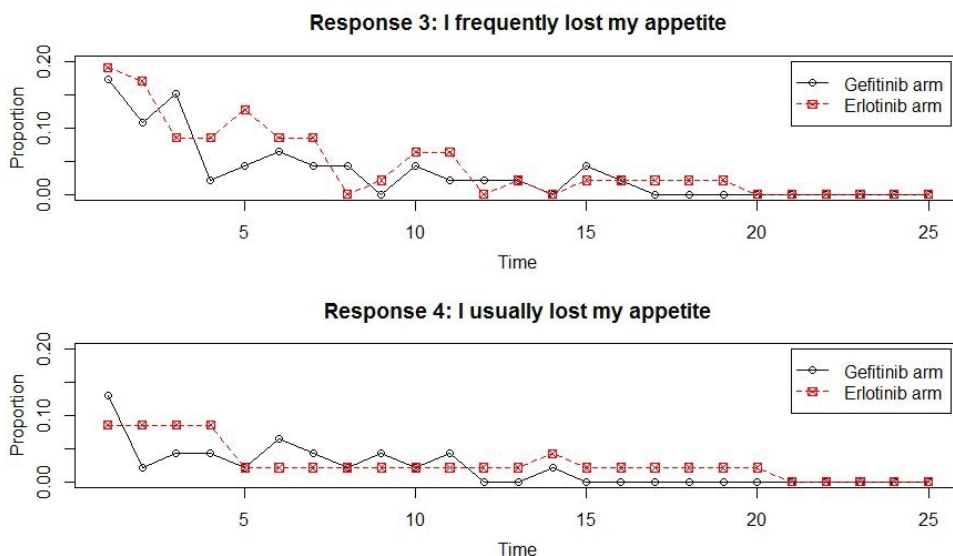


Figure 1: Marginal proportions for category 3 (I frequently lost my appetite) and 4 (I usually lost my appetite) given previous response under two arms (gefitinib and erlotinib). Time = 1 (baseline week), 2 (4 weeks), . . . , 25 (92 weeks).

progression-free survival (PFS) was 4.9 months (95% CI, 1.3–8.5) in gefitinib and 3.1 months (95% CI, 0.0–6.4) in erlotinib. However, there was no significant difference of the RR in the two arms. In this paper, it was of interest to see if there was a negative impact of treatments on patients quality of life (QOL). It is common that health-related QOL was evaluated with the use of a cancer-specific, 30-item score questionnaire developed by the European Organization for Research and Treatment of Cancer (EORTC). We compare the QOL data on the two treatments. The patients’ QOLs were measured at baseline, on day 1 of each subsequent 28-day cycle (4 weeks) and at the end of the study (week 0, 4, 8, . . . , 92). We concentrate on loss of appetite (Appetite) one of the QOL measures. Lee and Yoo (2014) also analyzed appetite outcomes, which were dichotomized. In this paper, we considered original scale (ordinal scale). Appetite was measured on a 4 point ordinal scale, 1: I always did not lose my appetite; 2: I sometimes lost my appetite; 3: I frequently lost my appetite; 4: I usually lost my appetite. Figure 1 presents respectively marginal proportions for category 3 (I frequently lost my appetite) and 4 (I usually lost my appetite) for two arms. The marginal proportions of appetite loss decreased as visit number increased and the two arms had similar pattern.

Most of patients dropped out of the study during the follow-up period. Missing data mechanism can be classified as missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR) (Daniels and Hogan, 2008). In this paper, the estimation of model parameters is based on Bayesian approach and we assume that the missing data are MAR in our analysis.

3. Bayesian cumulative logit random effects model for longitudinal ordinal data

We first present a cumulative logit random effects model for longitudinal ordinal data. Then we propose models for the random effects precision/covariance matrix in the cumulative logit model using MCD and MACD.

3.1. Cumulative logit random effects model

We first present the cumulative logit random effects model for longitudinal ordinal data. Let Y_{it} be the K -category ordinal response for subject i ($i = 1, \dots, N$) at time t ($t = 1, \dots, n_i$; $n_i \leq T$) and let x_{it} be the corresponding vector of covariates. We assume that each Y_{it} is conditionally independent given random effects b_{it} and the responses for different subjects are independent, and the regression model is given by

$$\log \frac{P(Y_{it} \leq k | b_{it}, x_{it})}{P(Y_{it} > k | b_{it}, x_{it})} = \beta_{0k} + x_{it}^T \beta + b_{it}, \quad (3.1)$$

where β_{0k} for $k = 1, \dots, K-1$ are intercepts that satisfy the monotonicity property (strictly increasing in k), x_{it} is a $p \times 1$ vector of covariates, β is a $p \times 1$ vector of regression coefficient. We assume

$$b_i = (b_{i1}, \dots, b_{in_i})^T \sim N(0, \Sigma_i), \quad (3.2)$$

where b_i is a vector of random effects values for subject i and Σ_i is the random effects covariance matrix which is decomposed into generalized time series parameters (GTSPs) and the IVs using MCD or MACD.

Decompositions provide parameters that can easily be modeled, have a sensible interpretation, and allow for simple computation (Daniels and Zhao, 2003; Lee *et al.*, 2012). The GTSPs consist of two types of parameters, GARP for the AR structure of Σ_i and GMAP for the MA structure of Σ_i , respectively. In addition these parameters can easily be modeled without the concern of the covariance matrix being positive definite.

3.2. Modeling of random effects precision matrix

In multivariate normal distribution, modeling of the precision matrix is easier than that of the covariance matrix. We now describe MCD for the random effects precision matrix. We assume that for $t = 2, \dots, n_i$,

$$b_i = (b_{i1}, \dots, b_{in_i})^T \sim N(0, \Sigma_i), \quad (3.3)$$

$$b_{i1} = e_{i1}, \quad (3.4)$$

$$b_{it} = \sum_{j=1}^{t-1} \phi_{itj} b_{ij} + e_{it},$$

where $e_{it} \stackrel{indep}{\sim} N(0, \sigma_{it}^2)$ with $\text{Var}(e_i) = \text{diag}(\sigma_{i1}^2, \dots, \sigma_{in_i}^2) = D_i$. Then we reexpress (3.3) and (3.4) in matrix form as

$$T_i b_i = e_i, \quad (3.5)$$

where $b_i = (b_{i1}, b_{i2}, \dots, b_{in_i})^T$, $e_i = (e_{i1}, e_{i2}, \dots, e_{in_i})^T$ and

$$T_i = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi_{i21} & 1 & 0 & \cdots & 0 \\ -\phi_{i31} & -\phi_{i32} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\phi_{in_i1} & -\phi_{in_i2} & -\phi_{in_i3} & \cdots & 1 \end{pmatrix}.$$

Then the covariance matrix is given by

$$T_i \Sigma_i T_i^T = \text{Var}(e_i) = D_i. \quad (3.6)$$

The GARPs and the IVs are represented by ϕ_{itj} and σ_{it}^2 , respectively.

Proposition 1. Σ_i in (3.6) is positive definite when IVs are positive.

Proof: Since T_i is invertible, $\Sigma_i = T_i^{-1} D_i T_i^{-T}$. Now let x be an arbitrary nonzero vector. Then $x^T \Sigma_i x = x T_i^{-1} D_i T_i^{-T} x$. For $z = T_i^{-T} x$, $x^T \Sigma_i x = z^T D_i z$. Since T_i is invertible and $\sigma_{it}^2 > 0$ for $t = 1, \dots, n_i$, Σ_i is positive definite. \square

Equation (3.6) is reexpressed as follows

$$\Sigma_i^{-1} = T_i^T D_i^{-1} T_i. \quad (3.7)$$

We note from (3.7) that the random effects precision matrix is directly decomposed into the GARPs and the IVs. They can be modeled using time and/or subject-specific covariate vectors w_{itj} and h_{it} by setting

$$\phi_{itj} = w_{itj}^T \gamma, \quad \log \sigma_{it}^2 = h_{it}^T \lambda, \quad (3.8)$$

where γ is a $a \times 1$ vector of unknown dependence parameters and λ is a $b \times 1$ vector of unknown variance parameters (Daniels and Zhao, 2003; Lee, 2013; Lee *et al.*, 2012; Pourahmadi, 2000; Pourahmadi and Daniels, 2002). The GARPs/IVs parameters are modeled as functions of subject-specific covariates by design vectors w_{itj} and h_{it} . For example, $w_{itj}^T = (1_{(t-j=1)}, 1_{(t-j=1)} \times \text{gender}_i)$ corresponds to an AR(1) structure depending on the covariate, gender, where $1_{(\cdot)}$ is the indicator function. A remarkable feature of (3.8) is its flexibility in reducing the potentially high-dimensional and constrained parameters of the precision matrix Σ_i^{-1} to $a + b$ unconstrained parameters γ and λ (Pourahmadi, 1999). As Proposition 1 presents, Σ_i is positive definite when IVs are positive.

3.3. Modeling of random effects covariance matrix

MCD in Subsection 3.2 is used to model the random effects precision matrix. However, the covariance matrix, rather than the precision matrix, is typically of interest. Now we present the modeling of the random effects covariance matrix in the cumulative random effects model. We again assume that $b_i = (b_{i1}, \dots, b_{in_i})^T \sim N(0, \Sigma_i)$ and

$$b_{i1} = e_{i1}, \quad (3.9)$$

$$b_{it} = \sum_{j=1}^{t-1} l_{itj} e_{ij} + e_{it}, \quad \text{for } t = 2, \dots, n_i, \quad (3.10)$$

where $e_i \stackrel{\text{indep}}{\sim} N(0, D_i)$ for $e_i = (e_{i1}, e_{i2}, \dots, e_{in_i})^T$. By setting $L_i = T_i^{-1}$ we restate (3.9) and (3.10) in matrix form as

$$b_i = L_i e_i,$$

where

$$L_i = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ l_{i21} & 1 & 0 & \cdots & 0 \\ l_{i31} & l_{i32} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{in_1} & l_{in_2} & l_{in_3} & \cdots & 1 \end{pmatrix}.$$

Here l_{itj} is called the GMAPs. The random effects covariance matrix Σ_i is

$$\Sigma_i = L_i D_i L_i^T. \quad (3.11)$$

We note that the random effects covariance matrix is decomposed into the GMAPs and the IVs. Rothman *et al.* (2010) described the relationship between L_i and T_i . Therefore, we can write $\Sigma_i = T_i^{-1} D_i T_i^{-T}$ which is equivalent to (3.6).

Proposition 2. Σ_i in (3.11) is positive definite when IVs are positive.

Proof: Let x be an arbitrary nonzero vector. Then $x^T \Sigma_i x = x L_i D_i L_i^T x$. For $z = L_i^T x$, $x^T \Sigma_i x = z^T D_i z$. Since L_i is invertible and $\sigma_{it}^2 > 0$ for $t = 1, \dots, n_i$, Σ_i is positive definite. \square

The parameters, GMAPs and IVs can be modeled z_{itj} and h_{it} by setting the model

$$l_{itj} = z_{itj}^T \gamma, \quad \log \sigma_{it}^2 = h_{it}^T \lambda,$$

where γ and λ are $a \times 1$ and $b \times 1$ vectors of unknown dependence and variance parameters, respectively. Note that design vectors z_{itj} and h_{it} are used to model the GMAP/IV parameters as functions of subject-specific covariates. GMAP/IV parametrization have several advantages. First, the parameters have a nice interpretation because of the linear combination of covariates in GMAPs and IVs. Second, the covariance matrix Σ_i can be heterogeneous depending on covariates when it is calculated using GMAPs and IVs.

3.4. Bayesian approach

In this subsection, we present a Bayesian approach for our proposed models. While the choice of prior distribution has been studied for linear regression model, there has been less work for GLMs. Propriety of the posterior distribution for logistic regression model using noninformative prior was presented in Chen and Shao (2001) and Ibrahim and Laud (1991). In our model, we use proper prior distribution instead of noninformative prior to guarantee the propriety of posterior distribution. The prior distributions for $\beta_{01}, \dots, \beta_{0(K-1)}, \beta, \gamma$, and λ in the model with the random effects precision/covariance matrix are given by

$$\beta \sim N(0, \sigma_\beta^2 I), \quad (3.12)$$

$$\beta_{01} \sim N(0, \sigma_{\beta_{01}}^2), \quad (3.13)$$

$$\beta_{0k} \sim N(0, \sigma_{\beta_{0k}}^2) 1_{(\beta_{0,k-1}, \infty)}, \quad \text{for } k = 2, \dots, K-1, \quad (3.14)$$

$$\gamma \sim N(0, \sigma_\gamma^2 I), \quad (3.15)$$

$$\lambda \sim N(0, \sigma_\lambda^2 I), \quad (3.16)$$

where I is the identity matrix, $\sigma_{\beta_{01}}^2, \sigma_{\beta_{02}}^2, \dots, \sigma_{\beta_{0,K-1}}^2, \sigma_{\beta}^2, \sigma_{\gamma}^2$, and σ_{λ}^2 are in general large to be noninformative (Daniels and Zhao, 2003). Since the intercepts β_0 have a constraint ($\beta_{01} < \dots < \beta_{0,K-1}$), we used the truncated normal distributions (SAS manual, http://support.sas.com/rnd/app/examples/stat/BayesMulti/new_example/).

From the sampling distribution (3.1) and prior distribution (3.12)–(3.16), we have the joint distribution given by

$$P(y, b, \beta, \beta_0, \gamma, \lambda) \propto \left[\prod_{i=1}^N \left\{ \prod_{t=1}^{n_i} \prod_{k=1}^K \pi_{itk}(b_i)^{y_{itk}} \right\} \phi(b_i | \gamma, \lambda) \right] \\ \times \phi(\beta_0 | 0, \sigma_{\beta_0}^2) 1_{(\beta_{01} < \dots < \beta_{0,K-1})} \phi(\beta | 0, \sigma_{\beta}^2) \phi(\gamma | 0, \sigma_{\gamma}^2) \phi(\lambda | 0, \sigma_{\lambda}^2),$$

where $\pi_{itk}(b_i) = P_{itk}^c(b_i) - P_{it(k-1)}^c(b_i)$ with $P_{itk}^c(b_i) = P(Y_{it} \leq k | b_{it}, x_{it})$ ($P_{it0}^c(b_{it}) = 0, P_{itK}^c(b_{it}) = 1$) for $k = 1, \dots, K$ and $\phi(\cdot)$ is the multivariate normal distribution function.

Since full conditionals are not closed forms (see Appendix), the parameters are drawn using a random walk Metropolis-Hastings algorithm. In practice, posterior computation proceeds via Markov chain Monte Carlo (MCMC) which can be easily implemented by WinBUGS (<http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>). The truncated normal prior distributions for β_{0k} 's are implemented using WinBUGS development interface (WBDev; Spiegelhalter *et al.*, 2003). The MCMC algorithm simulates direct draws from the above full conditionals iteratively until convergence is achieved.

3.5. Model selection

In Bayesian modeling, there are two common model selection criteria, posterior predictive loss (Gelfand and Ghosh, 1998) and Deviance Information Criterion (DIC) (Spiegelhalter *et al.*, 2002). They both take into account goodness of fit while penalizing models for overfitting (complexity penalty) (Daniels and Hogan, 2008). The posterior predictive loss quantifies the fit of the model by comparing features of the (model-based) posterior predictive distribution to equivalent features of the observed data. The comparison is based on a user-chosen loss function such as squared error loss. We use DIC for model selection in this paper since it is not easy to choose a proper loss function in categorical data.

The DIC is similar to Akaike Information Criterion (AIC) and is computed as the mean deviance minus the deviance evaluated at the posterior mean. Let θ be the vector of all parameters in our proposed models. The DIC is given by

$$\text{DIC} = \overline{\text{Dev}(\theta)} + p_D \\ = D(\tilde{\theta}) + 2p_D \\ = 2\overline{\text{Dev}(\theta)} - D(\tilde{\theta}),$$

where $\text{Dev}(\theta) = -2 \log L(\theta|y)$ and $p_D = \overline{\text{Dev}(\theta)} - D(\tilde{\theta})$.

Here $L(\theta|y)$ is the likelihood of $y = (y_1, \dots, y_N)^T$, and $\overline{\text{Dev}(\theta)}$ is the posterior mean of deviance and $D(\tilde{\theta}) = \text{Dev}[E_{\theta}(\theta|y)]$. Note that $\overline{\text{Dev}(\theta)}$, which is evaluated at the posterior mean of the parameters, is for a goodness-of-fit term and p_D is the effective number of parameters. For random effects models, the dimension of the parameter space is less clear and depends on the degree of heterogeneity between subjects.

4. Simulation study

We conducted small simulation studies to examine the properties of the proposed model to illustrate the performance of our algorithm.

Study 1. We simulated longitudinal data under our proposed model (3.1) and (3.2) with a precision matrix. The model was specified as

$$\log \frac{P(Y_{it} \leq k | b_{it}, x_{it})}{P(Y_{it} > k | b_{it}, x_{it})} = \beta_{0k} + \beta_1 \text{Group}_i + \beta_2 \text{Time}_{ij} + \beta_3 \text{Group}_i \times \text{Time}_{ij} + b_{it}, \quad (4.1)$$

for $j = 1, \dots, 10$ where $(\beta_{01}, \beta_{02}, \beta_{03}, \beta_1, \beta_2, \beta_3) = (-1.4, 0.1, 1.5, 0.1, -0.1, 0.1)$, $\text{Time}_{it} \sim N(0, 1)$ for $t = 1, \dots, 10$, Group_i equals 0 or 1 with an approximately equal sample size per group, $\phi_{ijl} = 0.9 \times 1_{|j-l|=1}$, and $\log \sigma_{ij}^2 = 0.1$.

We generated 30 simulated data sets each with a sample size of 500. Then we fit 3 models with the simulated data using Bayesian methodology. For each model, we use the same true mean but different covariance structures. Model 1 is the true model; Model 2 has a structure of the covariance matrix with $l_{ijl} = \gamma_0 \times I_{|j-l|=1}$ and $\log \sigma_{ij}^2 = \lambda_0$; Model 3 has a random intercept model with a scalar variance, $\log \sigma_{ij}^2 = \lambda_0$. For missingness, we specified the following MAR dropout model,

$$\text{logit}P(\text{dropout} = t | \text{dropout} \geq t) = -2.0 + 0.1Y_{it-1}.$$

Based on this specification, the observed dropout rates were approximately 50%.

Table 1 presents the average of posterior means of estimates and the percent relative biases of the parameters. In the presence of MAR missingness, the estimates in Models 1 and 2 had smaller biases than those in Model 3. In particular, the percent relative biases of β_0 s show that Models 1 and 2 fit better than Model 3. It indicates that ignoring the serial correlation structure can induce large biases in mean parameters. The biases in Models 1 and 2 were relatively small since the true covariance structure was assumed to be covariate-independent random effects covariance matrix (i.e., homoscedastic). We also calculated the estimated covariance matrix using the mean of 30 fitted values of γ and λ . Then we calculated the sum of absolute difference (SAD) between the estimated and true covariance matrices. The SAD for Model 1 was the smallest among the three models (as expected).

Study 2. We also generated 30 longitudinal data with a sample size of 500 using Model (4.1) with a random effects covariance matrix by setting $\phi_{ijl} = 0.9 \times 1_{|j-l|=1}$ and $\log \sigma_{ij}^2 = 0.2 + 0.2\text{Group}_i$. That is, the random effects covariance matrix was heteroscedastic (covariate-dependent) (Heagerty and Kurland, 2001). Similar to Study 1, we fit three models. Model 1 is the true model (with heteroscedastic AR(1) structure of the covariance matrix); Model 2 has a structure of the homogeneous AR(1) covariance matrix with $\phi_{ijl} = \gamma_0 \times I_{|j-l|=1}$ and $\log \sigma_{ij}^2 = \lambda_0$; Model 3 has a structure of the homogeneous MA(1) covariance matrix with $l_{ijl} = \gamma_0 \times I_{|j-l|=1}$ and $\log \sigma_{ij}^2 = \lambda_0$.

Table 2 presents the average of posterior means of estimates and the percent relative biases of the parameters. The estimates in Model 1 had smaller biases than those in Models 2 and 3. It indicates that ignoring heteroscedastic structure can induce biases in mean parameters. The SAD for Model 1 was the smallest among the three models (as expected). We also calculated average DIC values for three models using the 30 longitudinal data. As we expected, the DIC for Model 1 was the smallest among the three models.

Table 1: Simulation results for parameter estimates in Study 1 (bias of Models 1–3 for sample sizes of 500 and $T = 10$ for 30 simulated data sets under mis-specification of dependence).

	Model 1		Model 2		Model 3	
	Mean (SD)	RB	Mean (SD)	RB	Mean (SD)	RB
β_{01} (-1.4)	-1.422 (0.099)	1.6	-1.407 (0.103)	0.5	-1.155 (0.085)	-17.5
β_{02} (0.1)	0.096 (0.092)	-4.0	0.104 (0.096)	4.0	0.081 (0.083)	-19.0
β_{03} (1.5)	1.517 (0.103)	1.1	1.509 (0.104)	0.6	1.227 (0.086)	-18.2
β_1 (0.1)	0.109 (0.054)	9.0	0.105 (0.053)	5.0	0.086 (0.045)	-14.0
β_2 (-0.1)	-0.109 (0.132)	9.0	-0.114 (0.131)	14.0	-0.092 (0.116)	- 8.0
β_3 (0.1)	0.088 (0.073)	-12.0	0.093 (0.075)	-7.0	0.077 (0.064)	-23.0
γ_0 (0.9)	0.897 (0.026)		0.899 (0.025)			
λ_0 (0.1)	0.131 (0.167)		0.109 (0.154)		0.212 (0.098)	
SAD	16.25		101.89		127.76	

Displayed is the average posterior mean of regression coefficient estimate (Mean), the percent relative bias (RB) ($(\bar{\beta} - \beta)/\beta \times 100$), the average posterior standard deviation (SD) of 30 estimates, and sum of absolute difference (SAD) between the estimated and true covariance matrices for Models 1–3.

5. Analysis of lung cancer data

We used the approach proposed in Section 3 to analyze the QOL data from Section 2. In our analysis, we used 93 patients without missing data at baseline. To examine treatment differences in appetite levels, we included each type of treatment (Arm = 1 for Gefitinib, 0 for Erlotinib) and visit numbers (Time = 0.0, 0.1, . . . , 2.4) re-scaled. We assumed the missing responses (mostly due to dropout) were missing at random in our initial analysis.

5.1. Models fit

We fit eight models for Σ_i under an assumption of ignorable dropout (Table 3). Models 1 and 2 are cumulative logit random effects models with a random effects precision matrix having $\phi_{ijl} = \gamma_0 \times I_{|j-l|=1}$. The precision covariance matrix for Model 1 is homogeneous and that for Model 2 is heterogeneous depending on arm. Models 3 and 4 are cumulative logit random effects models with a random effects precision matrix having $\phi_{ijl} = \gamma_0 I_{(|l-j|=1)} + \gamma_2 I_{(|l-j|=2)}$. Model 3 has a homogeneous precision matrix and Model 4 has a precision matrix depending on arm. Models 5–8 are cumulative logit random effects models with a random effects covariance matrix having similar structures to Models 1–4. Table 3 shows detailed descriptions of Models 1–8.

For the estimation of the eight models, MCMC was implemented using WinBUGS 1.4.3 and R using the function R2WinBUGS (Sturtz *et al.*, 2005). The WinBUGS code is available upon request. The posterior means were calculated with a sample size of 50,000, burn-in period of 3,000, and thin of 5.

The posterior means, 95% credible intervals, and the DIC values for Models 1–8 are given in Table 4. We compare Models 1–8 using the DICs. The DIC values for Models 1–8 are 1305.37, 1304.86,

Table 2: Simulation results for parameter estimates in Study 2 (bias of Models 1–3 for sample sizes of 500 and $T = 10$ for 30 simulated data sets under mis-specification of dependence).

	Model 1		Model 2		Model 3	
	Mean (SD)	RB	Mean (SD)	RB	Mean (SD)	RB
β_{01} (-1.4)	-1.439 (0.127)	2.8	-1.379 (0.096)	- 1.5	-1.378 (0.097)	- 1.4
β_{02} (0.1)	0.082 (0.101)	-18.0	0.079 (0.093)	-21.0	0.080 (0.095)	-20.0
β_{03} (1.5)	1.512 (0.116)	0.8	1.449 (0.097)	- 3.4	1.449 (0.098)	- 3.4
β_1 (0.1)	0.110 (0.054)	10.0	0.113 (0.052)	13.0	0.113 (0.053)	13.0
β_2 (-0.1)	-0.079 (0.139)	-21.0	-0.073 (0.129)	-27.0	-0.073 (0.131)	-27.0
β_3 (0.1)	0.087 (0.078)	-13.0	0.080 (0.074)	-20.0	0.080 (0.074)	-20.0
γ_0 (0.9)	0.892 (0.029)		0.922 (0.016)		0.922 (0.016)	
λ_0 (0.1)	0.173 (0.204)		0.000 (0.334)		0.000 (0.317)	
λ_1 (0.2)	0.161 (0.140)					
SAD	17.862		25.955		126.649	
DIC	9402.041		9441.233		9441.065	

Displayed is the average posterior mean of regression coefficient estimate (Mean), the percent relative bias (RB) $((\hat{\beta} - \beta)/\beta \times 100)$, the average posterior standard deviation (SD) of 30 estimates, and sum of absolute difference (SAD) between the estimated and true covariance matrices for Models 1–3.

Table 3: Models fit with w_{itj} , z_{itj} and h_{it}

Model 1	$\phi_{itj} = \gamma_0 I_{(t-j)=1}$	$\log \sigma_{it}^2 = \lambda_0$
Model 2	$\phi_{itj} = \gamma_0 I_{(t-j)=1} + \gamma_1 I_{(t-j)=1} \text{Arm}_i$	$\log \sigma_{it}^2 = \lambda_0 + \lambda_1 \text{Arm}_i$
Model 3	$\phi_{itj} = \gamma_0 I_{(t-j)=1} + \gamma_2 I_{(t-j)=2}$	$\log \sigma_{it}^2 = \lambda_0$
Model 4	$\phi_{itj} = \gamma_0 I_{(t-j)=1} + \gamma_1 I_{(t-j)=1} \text{Arm}_i + \gamma_2 I_{(t-j)=2} + \gamma_3 I_{(t-j)=2} \text{Arm}_i$	$\log \sigma_{it}^2 = \lambda_0 + \lambda_1 \text{Arm}_i$
Model 5	$l_{itj} = \gamma_0 I_{(t-j)=1}$	$\log \sigma_{it}^2 = \lambda_0$
Model 6	$l_{itj} = \gamma_0 I_{(t-j)=1} + \gamma_1 I_{(t-j)=1} \text{Arm}_i$	$\log \sigma_{it}^2 = \lambda_0 + \lambda_1 \text{Arm}_i$
Model 7	$l_{itj} = \gamma_0 I_{(t-j)=1} + \gamma_2 I_{(t-j)=2}$	$\log \sigma_{it}^2 = \lambda_0$
Model 8	$l_{itj} = \gamma_0 I_{(t-j)=1} + \gamma_1 I_{(t-j)=1} \text{Arm}_i + \gamma_2 I_{(t-j)=2} + \gamma_3 I_{(t-j)=2} \text{Arm}_i$	$\log \sigma_{it}^2 = \lambda_0 + \lambda_1 \text{Arm}_i$

1212.49, 1215.93, 1378.58, 1378.99, 1334.51 and 1338.42, respectively. It indicates that Model 3 substantially is the best out of all the eight models. However, Models 3 and 4 had comparable DIC values. The posterior means of estimates for fixed effect parameters were similar in the two models.

We now focus on Models 3 by parsimonious rule. Remember Appetite was measured on a 4 point ordinal scale (1: I always did not lose my appetite; 2: I sometimes lost my appetite; 3: I frequently lost my appetite; 4: I usually lost my appetite). In Model 3, the 95% credible intervals for coefficients of Time and interaction of Time and Arm did not include zero indicating that patients' QOL were different between two arms. In erlotinib, the log odds of the conditional posterior probability of losing appetite decreased by 0.1721 ($0.1721 = (1.721 \times 0.1)$) as visit number increased by 1. In gefitinib, the log odds of the conditional posterior probability of losing appetite increased by 0.0647 ($-0.0647 = (1.721 - 2.368) \times 0.1$) as visit number increase by 1.

Table 4: Posterior means for Bayesian cumulative logit random effects model (95% credible intervals in the parentheses)

	Modeling of random effects precision matrix			
	Model 1 (Constant)	Model 2 (Arm)	Model 3 (Constant)	Model 4 (Arm)
Intercept: β_0				
β_{01}	-2.121*(-3.525, -0.880)	-2.121*(-3.451, -0.845)	-2.623*(-4.717, -0.906)	-2.617*(-4.655, -1.003)
β_{02}	1.392*(0.108, 2.665)	1.417*(0.184, 2.703)	2.015*(0.344, 4.139)	1.973*(0.422, 3.938)
β_{03}	3.713*(2.373, 5.178)	3.761*(2.431, 5.215)	5.035*(2.980, 8.301)	4.968*(3.001, 7.892)
Fixed parameters: β				
Time/10	1.425*(0.288, 2.648)	1.442*(0.340, 2.590)	1.721*(0.206, 3.473)	1.737*(0.356, 3.371)
Arm	1.786*(0.088, 3.539)	1.810*(0.060, 3.611)	2.111 (-0.142, 4.602)	2.148 (-0.084, 4.578)
Time/10×Arm	-1.950*(-3.646, -0.299)	-2.038*(-3.744, -0.386)	-2.368*(-4.820, -0.230)	-2.449*(-4.828, -0.307)
GARP: γ				
γ_0 (AR(1))	0.935*(0.868, 0.996)	0.894*(0.79, 0.983)	0.605*(0.457, 0.796)	0.564*(0.377, 0.799)
γ_1 (AR(1)×Arm)		0.073 (-0.043, 0.195)		0.100 (-0.174, 0.374)
γ_2 (AR(2))			0.323*(0.142, 0.476)	0.320*(0.093, 0.519)
γ_3 (AR(2)×Arm)				-0.015*(-0.311, 0.277)
IV parameters: λ				
λ_0	1.084*(0.489, 1.734)	1.026*(0.324, 1.759)	2.079*(1.118, 3.268)	1.979*(1.010, 3.096)
λ_1 (Arm)		0.185 (-0.426, 0.791)		0.155 (-0.366, 0.675)
Dev(θ)	1027.340	1021.860	824.203	830.097
p_D	278.022	283.003	388.282	385.830
DIC	1305.370	1304.860	1212.490	1215.930
	Modeling of random effects covariance matrix			
	Model 5 (Constant)	Model 6 (Arm)	Model 7 (Constant)	Model 8 (Arm)
Intercept: β_0				
β_{01}	-1.768*(-2.849, -0.775)	-1.691*(-2.736, -0.687)	-1.935*(-3.203, -0.79)	-1.770*(-2.947, -0.659)
β_{02}	1.175*(0.150, 2.191)	1.258*(0.252, 2.262)	1.350*(0.182, 2.552)	1.467*(0.394, 2.588)
β_{03}	3.158*(2.095, 4.254)	3.248*(2.200, 4.336)	3.532*(2.279, 4.931)	3.623*(2.477, 4.895)
Fixed parameters: β				
Time/10	1.188*(0.268, 2.123)	1.132*(0.278, 2.037)	1.292*(0.257, 2.409)	1.172*(0.213, 2.186)
Arm	1.502*(0.172, 2.883)	1.432*(0.062, 2.864)	1.613*(0.023, 3.243)	1.459 (-0.073, 3.003)
Time/10×Arm	-1.623*(-2.947, -0.336)	-1.608*(-2.985, -0.312)	-1.798*(-3.312, -0.295)	-1.712*(-3.206, -0.258)
GMAP: η				
η_0 (AR(1))	0.923*(0.847, 0.989)	0.878*(0.759, 0.977)	0.474*(0.353, 0.637)	0.442*(0.286, 0.661)
η_1 (AR(1)×Arm)		0.076 (-0.055, 0.216)		0.098 (-0.163, 0.359)
η_2 (AR(2))			0.384*(0.251, 0.508)	0.374*(0.186, 0.545)
η_3 (AR(2)×Arm)				-0.002 (-0.261, 0.248)
IV parameters: λ				
λ_0	-0.456 (-1.009, 0.119)	-0.486 (-1.19, 0.189)	0.692 (-0.086, 1.537)	0.553 (-0.304, 1.409)
λ_1 (Arm)		0.133 (-0.668, 0.913)		0.119 (-0.692, 0.935)
Dev(θ)	1179.980	1177.340	1086.430	1096.390
p_D	198.597	201.648	248.082	242.025
DIC	1378.580	1378.990	1334.510	1338.420

GARP=generalized autoregressive parameter; AR=autoregressive; IV=innovation variance; DIC=Deviance Information Criterion. * indicates the 95% credible interval does not include zero.

In the log(IV), the intercept (λ_0) was above zero which indicates random variation between random effects. The posterior means for D were given by

$$E(D|y) = \text{diag}(7.996, 7.996, \dots, 7.996, 7.996).$$

In the GARP, the posterior mean of coefficient (γ_0) for lag 1 was positive and the credible interval was

above zero which implies a positive relationship between random effects with lag 1. The posterior mean of coefficient for lag 2 also indicates a positive relationship between random effects with lag 2.

5.2. Convergence check

Now we focus on the best model, Model 3 and we check the model convergence using both history plots in WinBUGS. Figure 2 presents history plots with two chains with different starting points. We observed that the lines of different chains mix and cross in trace plots; therefore, the convergence is ensured.

6. Conclusion

We proposed Bayesian cumulative random effects models for longitudinal ordinal data with the general random effects precision/covariance matrix using MCD or MACD, respectively. The decompositions factor the random effects precision/covariance matrix to GARPs/GMAPs and IVs. The precision (covariance) matrix can be heterogeneous depending on covariates using the regression modeling of GARP (GMAP)/IV, and the covariance matrix is positive definite when IVs are positive.

The proposed models were implemented using the Bayesian approach for parameter estimation. WinBUGS program was used to fit the Bayesian models. Analysis results show that the random effects covariance matrix was homogeneous with the AR(2) structure. It indicates that between-subject variation was homogeneous over subjects and that lag-1 and 2 correlations were positive with lag-1 and 2 correlations of 0.993 and 0.997, respectively. The posterior cumulative conditional probability of appetite decreases as visit increases for gefitinib. Similarly, the posterior cumulative probability decreases for erlotinib. Therefore, the QOL of patients was different between two arms over time. The codes for our proposed models are available upon request.

MCD or MACD can be extended to nominal random effects models for analysis of longitudinal nominal data. However, the random effects covariance matrix in nominal random effects models may be a Kronecker product of the correlation matrix for serial dependence and the covariance matrix among categories at the same time (Galecki, 1994; Lee *et al.*, 2011). Therefore, the two decompositions are not directly extended for the nominal random effects models.

Acknowledgement

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (KRF) funded by the Ministry of Education, Science and Technology (NRF-2014R1A1A2054997, NRF-2016R1D1A1B03930343).

We would like to thank Drs. Myung-Ju Ahn and Insuk Sohn of Samsung Medical Center in Korea for providing the data and for their help in data collection and clarifying some issues with data.

Appendix: Full conditionals

Full conditional posterior distributions are required to implement the Markov chain Monte Carlo (MCMC) algorithm as:

- For β

$$P(\beta|y, b, \beta_0, \gamma, \lambda) \propto \left[\prod_{i=1}^N \left\{ \prod_{t=1}^{n_i} \prod_{k=1}^K (P_{itk}^c(b_i) - P_{it(k-1)}^c(b_i))^{y_{itk}} \right\} \right] \exp \left(-\frac{1}{2\sigma_\beta^2} \beta^T \beta \right).$$

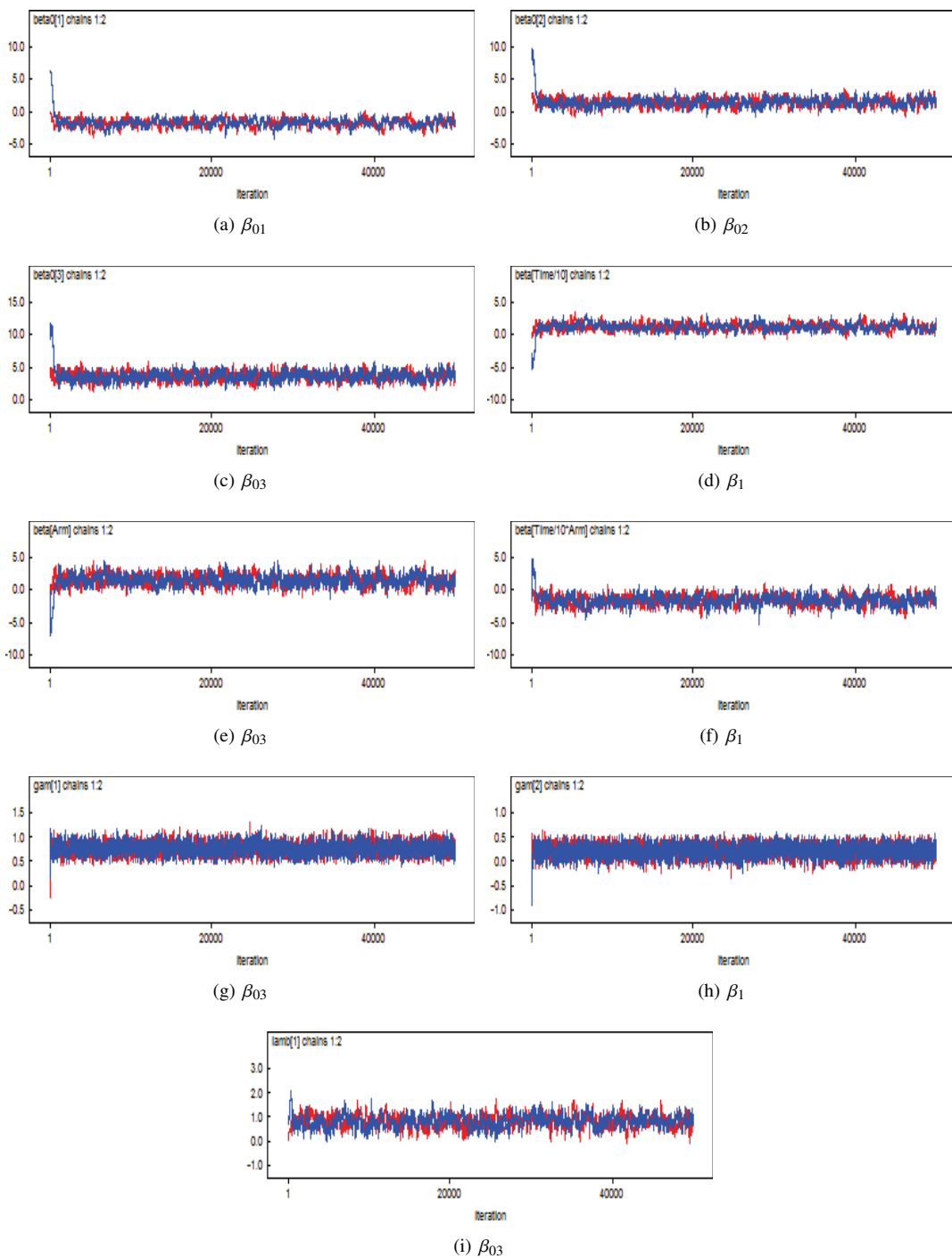


Figure 2: Trace plots of Model 3 in WinBUGS.

- For γ

$$P(\gamma|y, b, \beta, \beta_0, \lambda) \propto \exp\left(-\frac{1}{2} \sum_{i=1}^N b_i^T \Sigma_i^{-1} b_i\right) \exp\left(-\frac{1}{2\sigma_\gamma^2} \gamma^T \gamma\right).$$

- For λ

$$P(\lambda|y, b, \beta, \beta_0, \gamma) \propto \left\{ \prod_{i=1}^N \prod_{t=1}^{n_i} (\sigma_{it}^2)^{-\frac{1}{2}} \right\} \exp\left(-\frac{1}{2} \sum_{i=1}^N b_i^T \Sigma_i^{-1} b_i\right) \exp\left(-\frac{1}{2\sigma_\lambda^2} \lambda^T \lambda\right).$$

- For b_i ($i = 1, \dots, N$)

$$P(b_i|y, \beta, \beta_0, \gamma, \lambda) \propto \left\{ \prod_{t=1}^{n_i} \prod_{k=1}^K (P_{itk}^c(b_i) - P_{it(k-1)}^c(b_i))^{y_{itk}} \right\} \exp\left(-\frac{1}{2} \sum_{i=1}^N b_i^T \Sigma_i^{-1} b_i\right).$$

- For β_{01}

$$P(\beta_{01}|y, b, \beta, \gamma, \lambda) \propto \left[\prod_{i=1}^N \left\{ \prod_{t=1}^{n_i} (P_{it1}^c(b_i) - P_{it0}^c(b_i))^{y_{it1}} \right\} \right] \exp\left\{-\frac{1}{2\sigma_{\beta_{01}}^2} \beta_{01}^T \beta_{01}\right\}.$$

- For β_{0k} on $(\beta_{0,k-1}^\infty)$ ($k = 2, \dots, K-1$)

$$P(\beta_{0k}|y, b, \beta, \gamma, \lambda) \propto \left[\prod_{i=1}^N \left\{ \prod_{t=1}^{n_i} \prod_{k=2}^K (P_{itk}^c(b_i) - P_{it,k-1}^c(b_i))^{y_{itk}} \right\} \right] \exp\left\{-\frac{1}{2\sigma_{\beta_{0k}}^2} \beta_{0k}^T \beta_{0k}\right\}.$$

Note that the full conditionals are not closed forms.

References

- Breslow NE and Clayton DG (1993). Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association*, **88**, 9–25.
- Chen MH and Shao QM (2001). Propriety of posterior distribution for dichotomous quantal response models, *Proceedings of the American Mathematical Society*, **129**, 293–302.
- Daniels JM and Pourahmadi M (2002). Bayesian analysis of covariance matrices and dynamic models for longitudinal data, *Biometrika*, **89**, 553–566.
- Daniels JM and Zhao YD (2003). Modelling the random effects covariance matrix in longitudinal data, *Statistics in Medicine*, **22**, 1631–1647.
- Daniels MJ and Hogan JW (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*, Chapman & Hall/CRC, Boca Raton, FL.
- Diggle PJ, Heagerty P, Liang KY, and Zeger SL (2002). *Analysis of Longitudinal Data* (2nd ed), Oxford University Press, New York.
- Galecki AT (1994). General class of covariance structures for two or more repeated factors in longitudinal data analysis, *Communications in Statistics-Theory and Method*, **23**, 3105–3119.
- Gelfand AE and Ghosh SK (1998). Model choice: a minimum posterior predictive loss approach, *Biometrika*, **85**, 1–11.

- Gelman A and Rubin D (1992). Inference from iterative simulation using multiple sequences, *Statistical Science*, **7**, 457–511.
- Gibbons RD and Hedeker D (1997). Random effects probit and logistic regression models for three-level data, *Biometrics*, **53**, 1527–1537.
- Heagerty PJ and Kurland BF (2001). Misspecified maximum likelihood estimates and generalised linear mixed models, *Biometrika*, **88**, 973–985.
- Ibrahim JG and Laud PW (1991). On Bayesian analysis of generalized linear models using Jeffreys's prior, *Journal of the American Statistical Association*, **86**, 981–986.
- Kim J and Lee K (2015). Survey of models for random effects covariance matrix in generalized linear mixed model, *Korean Journal of Applied Statistics*, **28**, 211–219.
- Kim ST, Uhm JE, Lee J, Sun J, Sohn I, Kim SW, Jung S, Park YH, Ahn JS, Park K, and Ahn MJ (2012). Randomized phase II study of gefitinib versus erlotinib in patients with advanced non-small cell lung cancer who failed previous chemotherapy, *Lung Cancer*, **75**, 82–88.
- Lee K (2013). Bayesian modeling of random effects covariance matrix for generalized linear mixed models, *Communication for Statistical Applications and Methods*, **20**, 235–240.
- Lee K and Daniels MJ (2008). Marginalized models for longitudinal ordinal data with application to quality of life studies, *Statistics in Medicine*, **27**, 4359–4380.
- Lee K, Daniels MJ, and Joo Y (2013). Flexible marginalized models for bivariate longitudinal ordinal data, *Biostatistics*, **14**, 462–476.
- Lee K, Kang S, Liu X, and Seo D (2011). Likelihood-based approach for analysis of longitudinal nominal data using marginalized random effects models, *Journal of Applied Statistics*, **38**, 1577–1590.
- Lee K and Yoo JK (2014). Bayesian Cholesky factor models in random effects covariance matrix for generalized linear mixed models, *Computational Statistics & Data Analysis*, **80**, 111–116.
- Lee K, Lee J, Hagan J, and Yoo JK (2012). Modeling the random effects covariance matrix for the generalized linear mixed models, *Computational Statistics & Data Analysis*, **56**, 1545–1551.
- Liu I and Agresti A (2005). The analysis of ordered categorical data: an overview and a survey of recent developments, *Test*, **14**, 1–73.
- Litiere S, Alonso A, and Molenberghs G (2007). Type I and type II error under random-effects misspecification in generalized linear mixed models, *Biometrics*, **63**, 1038–1044.
- Litiere S, Alonso A, and Molenberghs G (2008). The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models, *Statistics in Medicine*, **27**, 3125–3144.
- McCullagh P (1980). Regression models for ordinal data, *Journal of the Royal Statistical Society Series B (Methodological)*, **42**, 109–142.
- Nooraee N, Molenberghs G, and van den Heuvel ER (2014). GEE for longitudinal ordinal data: comparing R-geepack, R-multgee, R-repolr, SAS-GENMOD, SPSS-GENLIN, *Computational Statistics & Data Analysis*, **77**, 70–83.
- Pan J and MacKenzie G (2003). On modelling mean-covariance structures in longitudinal studies, *Biometrika*, **90**, 239–244.
- Pan J and MacKenzie G (2006). Regression models for covariance structures in longitudinal studies, *Statistical Modelling*, **6**, 43–57.
- Pourahmadi M (1999). Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation, *Biometrika*, **86**, 677–690.
- Pourahmadi M (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix, *Biometrika*, **87**, 425–435.

- Pourahmadi M and Daniels MJ (2002). Dynamic conditionally linear mixed models for longitudinal data, *Biometrics*, **58**, 225–231.
- Rothman AJ, Levina E, and Zhu J (2010). A new approach to Cholesky-based covariance regularization in high dimensions, *Biometrika*, **97**, 539–550.
- Spiegelhalter D, Best N, Carlin B, and van der Linde A (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, **64**, 583–639.
- Spiegelhalter D, Thomas A, Best N, and Lunn D (2003). *WinBUGS Version 1.4 User Manual*, Medical Research Council Biostatistics Unit, Cambridge, UK.
- Sturtz S, Ligges U, and Gelman A (2005). R2WinBUGS: a package for running WinBUGS from R *Journal of Statistical Software*, **12**, 1–16.
- Zhang W and Leng C (2012). A moving average Cholesky factor model in covariance modelling for longitudinal data, *Biometrika*, **99**, 141–150.

Received October 18, 2016; Revised December 3, 2016; Accepted December 5, 2016