Bayesian Pattern Mixture Model for Longitudinal Binary Data with Nonignorable Missingness

Yujung Kyoung[a], Keunbaik Lee[1,a]

[a]Department of Statistics, Sungkyunkwan University, Korea

Abstract

In longitudinal studies missing data are common and require a complicated analysis. There are two popular modeling frameworks, pattern mixture model (PMM) and selection models (SM) to analyze the missing data. We focus on the PMM and we also propose Bayesian pattern mixture models using generalized linear mixed models (GLMMs) for longitudinal binary data. Sensitivity analysis is used under the missing not at random assumption.

Keywords: pattern mixture model, nonignorable missing, sensitivity analysis, binary data

1. Introduction

In longitudinal studies missing data are common and require a complicated analysis. Missing values occur as *dropouts* whenever subjects are terminated early from a longitudinal study for reasons outside the control of the investigator. In this thesis, we consider a particular type of dropout, known as monotone dropout. Assume that the full-data response vector is $Y_i = (Y_{i1}, \ldots, Y_{iT})^T$. If Y_{it} is missing, then $Y_{it+1}, \ldots, Y_{iT}$ are all missing.

Diggle and Kenward (1994) provide a characterization of informative dropout for repeated outcomes; under monotone dropout, these correspond to sequential versions of Rubin's (1976) general taxonomy for missing-data mechanisms. Identifiability is a serious concern when missingness is potentially nonignorable; specifically, models that incorporate nonignorable missingness can be identified only through untestable parameter restrictions or distributional assumptions (Little, 1995; Rubin, 1977; Scharfstein *et al.*, 1999). Lack of a priori knowledge about the true missing data mechanism is frequently handled using sensitivity analyses (Scharfstein *et al.*, 1999). Sensitivity analysis is an assessment of sensitivity of model-based inferences to assumptions that cannot be verified or checked with data (Daniels and Hogan, 2008).

Restrictions on missing data mechanism can be classified as missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR). MCAR occurs when data are missing for reasons completely unrelated to the observed or missing parts of Y, conditionally on X. MAR is independent of missing responses Y_{mis}, conditionally on observed response Y_{obs} and model covariates X. MNAR is dependent on some part of Y_{mis}, even after conditioning on Y_{obs} and X.

To analyze longitudinal data with informative missingness, there are two popular modeling frameworks, pattern mixture model (PMM) and selection models (SM), which are different in the way that

the joint distribution of the outcome and missing data process are factorized (Daniels and Hogan, 2008; Hogan and Laird, 1997; Kenward and Molenberghs, 1999; Little, 1995; Molenberghs and Kenward, 2007). Both PMMs and SMs derive their inferences based on the joint distribution $f(Y, R)$ of the outcome Y and the missing-data indicator R, but use different decomposition for $f(Y, R)$.

SMs partition $P(Y, R)$ as the product of $f(Y)$ and $f(R|Y)$ (Heckman, 1974). They require explicit modeling of the missing-data mechanism where the probability of missingness depend on observed and unobserved values. In parametric selection models sensitivity analysis is not proper because the all parameters are not easily factored into identified and nonidentified parameters. Selection models (although identifiable) should be approached with special caution in the context of MNAR models (Daniels and Hogan, 2008). However, PMMs express the joint distribution as the product of $P(Y|R)$ and $f(R)$ where the data are stratified by the missing data patterns with distinct parameters for each pattern (Rubin, 1977; Little and Rubin, 2002). The marginal estimates in PMM can be derived as a weighted average across pattern specific estimates (Little, 1995) or by using imputation (Demirtas and Schafer, 2003). PMMs are well suited better than SMs because PMMs are easily factored with extrapolation factorization that enables interpretable sensitivity analysis and the formulation of missing data assumptions. Therefore, we use pattern mixture model.

This paper also develops Bayesian pattern mixture models using generalized linear mixed models (GLMMs) for longitudinal binary data. We also use sensitivity analyses under the MNAR assumption. The GLMMs are frequently used when subject-specific effect is of interest (Breslow and Clayton, 1993). In the GLMMs, the effect of covariates on response is a specified conditional on random effects. We also consider different patterns for a random effects covariance matrix.

This paper is organized as follows. In Section 2, we provides description of a motivating example, the Metabolic Syndrome data. In Section 3, we propose the pattern mixture GLMMs for repeated binary data. Finally, conclusions and extensions are provided in Section 4.

2. Motivating Data: Metabolic Syndrome Data

The Korean Genomic Epidemiology Study (KoGES) monitors the development of metabolic syndrome for male and female Korean adults 39–69 years of age (Kim *et al.*, 2006). In KoGES 2310 participants were examined every two years for eight years to monitor the development of metabolic syndrome.

Metabolic syndrome is defined when three or more of the following five disorders are present: abdominal obesity (waist circumference > 90cm in men or > 80cm in women), high blood pressure (systolic BP levels > 130mmHg or diastolic BP levels > 85mmHg), high impaired fasting glucose (IFG > 110mg/dl), high triglyceridemia (TG > 150mg/dl), and low high-density lipoprotein cholesterol (HDL-C < 40mg/dl in men or < 50mg/dl in women). The primary outcome measure was whether or not the patient had metabolic syndrome. Demographic factors of the metabolic syndrome were sex, age, alcohol consumption, and smoking. According to previous literature on the metabolic syndrome, the effect of covariate smoking was significant (Lee *et al.*, 2013). Therefore, we only considered the covariate smoking in this paper and examine the influence of smoking on metabolic syndrome.

Table 1 presents summarizes proportion metabolic syndrome, general, and missing at each year, stratified by smoking group, and includes two commonly used estimates of overall metabolic syndrome patients rate. MS/(MS + G) denotes metabolic syndrome rate and (MS + M)/(MS + G + M) denotes metabolic syndrome rate after counting missing values as metabolic syndrome patients. Here MS is the number of people with metabolic syndrome, G is the number of people without metabolic syndrome, and M is the number of people who dropped out. In both groups, the proportions of

Table 1: Metabolic syndrome: number of people with metabolic syndrome, without metabolic syndrome, and missing at each year

Group		Year			
		2	4	6	8
Smoking ($n = 1894$)	Metabolic syndrome	440 (.23)	385 (.20)	307 (.16)	285(.15)
	General	1454 (.77)	1373 (.73)	1372 (.72)	1131 (.60)
	Missing	0	136 (.07)	215 (.11)	478 (.25)
	$\dfrac{\text{MS}}{\text{MS} + G}$	0.23	0.22	0.18	0.20
	$\dfrac{\text{MS} + M}{\text{MS} + G + M}$	0.23	0.28	0.28	0.40
Non-smoking ($n = 416$)	Metabolic syndrome	108 (.26)	31 (.07)	0	0
	General	308 (.74)	98 (.24)	0	0
	Missing	0	287 (.69)	416 (1.00)	416 (1.00)
	$\dfrac{\text{MS}}{\text{MS} + G}$	0.26	0.24	0	0
	$\dfrac{\text{MS} + M}{\text{MS} + G + M}$	0.26	0.76	1.00	1.00

metabolic syndrome decreased in age. However, all participants in non-smoking group dropped out after 4 years. It indicates that typical analysis under MAR cannot be used in this case.

3. Pattern Mixture GLMMs for Repeated Binary Data

In this section, we describe generalized linear mixed models (GLMMs) for longitudinal binary data. Then we propose pattern mixture models for dropout missingness using the GLMMs.

3.1. Generalized linear mixed models

Let Y_{it} be a binary response at time t ($t = 1, \ldots, n_i$) for subject i ($i = 1, \ldots, N$), x_{it} be the corresponding vector of covariates, and b_{it} be the vector of random effects for subject i. We assume that $Y_i = \{Y_{i1}, \ldots, Y_{in_i}\}$ are conditionally independent given random effect b_{it}. The regression model is given by

$$\text{logit } p_{it}(b_{it}) = x_{it}\beta + b_{it}, \tag{3.1}$$

where β is the $p \times 1$ vector of regression coefficient and $p_{it}(b_{it}) = P(Y_{it} = 1; b_{it})$. We assume that

$$b_i = \left(b_{i1}, \ldots, b_{in_i}\right)^T \sim N(0, \Sigma_i),$$

where Σ_i is the random effects covariance matrix for subject i.

3.2. PMMs for longitudinal binary data

We now propose pattern mixture models under the missing not at random assumption. The GLMMs are composed within pattern for longitudinal binary data. In the GLMMs, the effect of covariates on response is specified conditional on random effects (b_i). We assume full data consists of repeated binary response $Y_i = (Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})$ for subject i ($i = 1, \ldots, N$) scheduled at times $t_1, \ldots, t_4$ (assume $t_j = j$) and $R_i = (R_{i1}, R_{i2}, R_{i3}, R_{i4})^T$ be the observed data indicators (assuming monotone dropout), where $R_{ij} = 0$ corresponds to Y_{ij} being missing. We also let Z_i be smoking indicator where $Z_i = 1$ is for the smoker and 0 for nonsmoker. We now define S_i to be the number of observed responses and assume only dropouts. The follow-up time $S_i = S(R_i)$ is defined as the final time point at which data

are observed below:

$$S_i = \begin{cases} 1, & \text{if } R_i = (1,0,0,0), \\ 2, & \text{if } R_i = (1,1,0,0), \\ 3, & \text{if } R_i = (1,1,1,0), \\ 4, & \text{if } R_i = (1,1,1,1). \end{cases}$$

Our proposed model is given by

$$\text{logit}\left(p_{it}^{(s)}(b_{it})\right) = \beta_0^{(s)} + z_i\beta_1^{(s)} + b_{it}, \tag{3.2}$$

where $p_{it}^{(s)}(b_{it}) = p(Y_{it} = 1|b_{it}, S = s)$. Similar to the PMM in Daniels and Hogan (2008), we assume multinomial distribution for the dropout time, S,

$$S \sim \text{Multinomial}(\phi),$$

where $\phi = (\phi_1, \phi_2, \phi_3, \phi_4)$ and $\sum_s \phi_S = 1$. We also assume that

$$b_i = (b_{i1}, \dots, b_{i4})|S = s \sim N\left(0, \Sigma^{(s)}\right).$$

The distribution of the random effects b_i is reexpressed as:

$$b_1|S = 1 \sim N\left(0, \sigma_1^2\right), \tag{3.3}$$

$$b_1|S = 2 \sim N\left(0, \sigma_1^2\right), \tag{3.4}$$

$$b_1|S = 3 \sim N\left(0, \sigma_1^2\right), \tag{3.5}$$

$$b_1|S = 4 \sim N\left(0, \sigma_1^2\right), \tag{3.6}$$

$$\mathbf{b_2|b_1, S = 1} \sim \mathbf{N}\left(\alpha_1^{(1)}\mathbf{b_1}, \sigma_2^{*2}\right), \tag{3.7}$$

$$b_2|b_1, S \geq 2 \sim N\left(\alpha_1^{(\geq 2)}b_1, \sigma_2^2\right), \tag{3.8}$$

$$\mathbf{b_3|b_1, b_2, S = 1, 2} \sim \mathbf{N}\left(\alpha_2^{(1:2)}\mathbf{b_1} + \theta_1^{(1:2)}\mathbf{b_2}, \sigma_3^{*2}\right), \tag{3.9}$$

$$b_3|b_1, b_2, S = 3, 4 \sim N\left(\alpha_2^{(\geq 3)}b_1 + \theta_1^{(\geq 3)}b_2, \sigma_3^2\right), \tag{3.10}$$

$$\mathbf{b_4|b_1, b_2, b_3, S = 1, 2, 3} \sim \mathbf{N}\left(\alpha_3^{(1:3)}\mathbf{b_1} + \theta_2^{(1:3)}\mathbf{b_2} + \gamma_1^{(1:3)}\mathbf{b_3}, \sigma_4^{*2}\right), \tag{3.11}$$

$$b_4|b_1, b_2, b_3, S = 4 \sim N\left(\alpha_3^{(4)}b_1 + \theta_2^{(4)}b_2 + \gamma_1^{(4)}b_3, \sigma_4^2\right). \tag{3.12}$$

We note that variance in each pattern is same to reduce the number of parameters. Now the parameters in (3.3)–(3.12) are reparameterized in terms of sensitivity parameters and we define the functions $h(\xi_M, \Delta)$ as under MNAR:

$$\alpha_j^{(s)} = h_1(\xi_M, \Delta) = \alpha_j^{(\geq j)} + \Delta_\alpha, \quad j = 1, 2, 3, \tag{3.13}$$

$$\theta_j^{(s)} = h_2(\xi_M, \Delta) = \theta_j^{(\geq j)} + \Delta_\theta, \quad j = 1, 2, \tag{3.14}$$

$$\gamma_j^{(s)} = h_3(\xi_M, \Delta) = \gamma_j^{(\geq j)} + \Delta_\gamma, \quad j = 1, \tag{3.15}$$

$$\sigma_j^{*2} = h_4(\xi_M, \Delta) = \sigma_j^2\Delta_\sigma, \quad j = 2, 3, 4, \tag{3.16}$$

where $\boldsymbol{\Delta}$ captures the information about the missing data mechanism. In general, priors for components of Δ_α, Δ_θ, and Δ_γ will shift the posterior log odds away from their MAR values.

Under this parameterization of the PMM, the conditional variance parameters σ^2 from nonidentified distributions do not appear in the posterior distribution of the marginal mean parameters. Hence, priors for components of Δ_σ will not affect posterior inference about marginal means. This justifies that we consider one Δ_σ.

3.3. Prior distributions

Prior distribution for parameters can still be specified as

$$P(\xi_S, \xi_M, \boldsymbol{\Delta}) = P(\xi_S | \xi_M, \boldsymbol{\Delta}) P(\boldsymbol{\Delta} | \xi_M) P(\xi_M), \tag{3.17}$$

where

$$P(\xi_S | \xi_M, \boldsymbol{\Delta}) = I[\xi_S = h(\xi_M, \boldsymbol{\Delta})]. \tag{3.18}$$

Here ξ_S are the (nonidentified) sensitivity parameters in the full-data model, ξ_M are (identified) parameters indexing the implied observed model, and $\boldsymbol{\Delta}$ captures departures from MAR. The function h represents the missing data mechanism, and makes explicit how priors or assumptions are being used to infer the full-data model (Rubin, 1977). In our case, $h(\xi_M, \boldsymbol{\Delta})$ are specified in (3.13)–(3.16).

In equation (3.17), $P(\xi_M)$ is identified, $P(\boldsymbol{\Delta} | \xi_M)$ is a flat distribution in a Bayesian perspective and $P(\xi_S | \xi_M, \boldsymbol{\Delta})$ is an indicator given like equation (3.18). It is a point mass, and $P(\boldsymbol{\Delta} | \xi_M)$ reflects prior beliefs about departures from MAR as follows,

$$P(\Delta_\alpha, \Delta_\theta, \Delta_\gamma, \Delta_\sigma) = I\{\Delta_\alpha = \Delta_\theta = \Delta_\gamma = 0, \Delta_\sigma = 1\}.$$

The priors for $\boldsymbol{\Delta}$ and ξ_M as:

$$\beta_0^{(s)} \sim \text{Normal}\left(0, \sigma_{\beta_0}^2 I\right), \tag{3.19}$$

$$\beta_1^{(s)} \sim \text{Normal}\left(0, \sigma_{\beta_1}^2 I\right), \tag{3.20}$$

$$\alpha_j^{(s)} \sim \text{Normal}\left(0, \sigma_{\alpha_j}^2 I\right), \tag{3.21}$$

$$\theta_j^{(s)} \sim \text{Normal}\left(0, \sigma_{\theta_j}^2 I\right), \tag{3.22}$$

$$\gamma_1^{(s)} \sim \text{Normal}\left(0, \sigma_{\gamma_1}^2 I\right), \tag{3.23}$$

$$\sigma_j^2 \sim \text{Uniform}\left(0, a_{\sigma^2}\right), \tag{3.24}$$

$$\Delta_{\alpha \cdot \theta \cdot \gamma} \sim \text{Uniform}\left(c_1, c_2\right), \tag{3.25}$$

$$\Delta_\sigma \sim \text{Uniform}\left(0, c_3\right), \tag{3.26}$$

where $\sigma_{\beta_0}^2$, $\sigma_{\beta_1}^2$, $\sigma_{\alpha_j}^2$, $\sigma_{\theta_j}^2$, and $\sigma_{\gamma_j}^2$ are large value (for example, 100). a_{σ^2}, c_1, c_2, and c_3 are proper values.

3.4. Bayesian analysis via MCMC

To derive the likelihood function for the GLMM expressions (3.1) and (3.19)–(3.23) we let $\omega = (\beta_0, \beta_1, \omega_2)$ where $\omega_1 = (\alpha, \theta, \gamma, \sigma^2, \Delta_{\alpha \cdot \theta \cdot \gamma}, \Delta_\sigma)$. The joint distribution of sample in (3.2) and the ran-

dom effects is given by

$$L(\omega; y, b, s) = \prod_{i=1}^{N} p(y_i|b_i, s_i, \omega)p(b_i|s_i, \omega_1)p(s_i), \qquad (3.27)$$

where

$$p(y_i|b_i, s_i, \omega) = \prod_{t=1}^{4} \left(p_{it}^{(s_i)}\right)^{y_{it}} \left(1 - p_{it}^{(s_i)}\right)^{1-y_{it}},$$

$$p(b_i|s_i, \omega_1) = p(b_{i1}|s_i)p(b_{i2}|b_{i1}, s_i)p(b_{i3}|b_{i2}, b_{i1}, s_i)p(b_{i4}|b_{i3}, b_{i2}, b_{i1}, s_i). \qquad (3.28)$$

Note that the conditional distributions of the random effects b_i in (3.28) are given in (3.3)–(3.12). From the distribution (3.27) and prior distributions (3.19)–(3.26), the joint distribution is given by

$$p(y, b, s, \omega) \propto \left\{ \prod_{i=1}^{N} p(y_i|b_i, s_i, \omega)p(b_i|s_i, \omega_1)p(s_i) \right\} \pi(\beta_0)\pi(\beta_1)\pi(\alpha)\pi(\theta)\pi(\gamma)\pi\left(\sigma^2\right)\pi(\Delta_{\alpha\cdot\theta\cdot\gamma})\pi(\Delta_\sigma).$$

Full conditional posterior distributions are required to implement the MCMC algorithm (Gelman *et al.*, 2004) and are given as:

- For $\beta_0^{(s)}$, $s = 1, \ldots, 4$,

$$p\left(\beta_0^{(s)}|\omega_{-\beta_0^{(s)}}\right) \propto \left[\prod_{i\in A_s} \left\{ \prod_{t=1}^{4} \left(p_{it}^{(s)}\right)^{y_{it}} \left(1 - p_{it}^{(s)}\right)^{1-y_{it}} \right\} \right] \exp\left(-\frac{\beta_0^{(s)2}}{2\sigma_{\beta_0}^2}\right),$$

  where $A_s = \{i|\text{subject } i \text{ belongs to Pattern } s\}$.

- For $\beta^{(s)}$,

$$p\left(\beta_1^{(s)}|\omega_{-\beta^{(s)}}\right) \propto \left[\prod_{i\in A_s} \left\{ \prod_{t=1}^{4} \left(p_{it}^{(s)}\right)^{y_{it}} \left(1 - p_{it}^{(s)}\right)^{1-y_{it}} \right\} \right] \exp\left(-\frac{\beta^{(s)2}}{2\sigma_\beta^2}\right).$$

- For $\alpha_j^{(s)}$, $j = 1, 2, 3$,

$$p\left(\alpha^{(s)}|\omega_{-\alpha^{(s)}}\right) \propto \left\{ \prod_{i\in A_s} p(b_i|s_i = s, \omega) \right\} \exp\left(-\frac{\alpha^{(s)2}}{2\sigma_\alpha^2}\right).$$

- For $\theta_j^{(s)}$, $j = 1, 2$,

$$p\left(\theta^{(s)}|\omega_{-\theta^{(s)}}\right) \propto \left\{ \prod_{i\in A_s} p(b_i|s_i = s, \omega) \right\} \exp\left(-\frac{\theta^{(s)2}}{2\sigma_\theta^2}\right).$$

- For $\gamma_1^{(s)}$,

$$p\left(\gamma_1^{(s)}|\omega_{-\gamma_1^{(s)}}\right) \propto \left\{ \prod_{i\in A_s} p(b_i|s_i = s, \omega) \right\} \exp\left(-\frac{\gamma_1^{(s)2}}{2\sigma_\gamma^2}\right).$$

- For σ_j^2, $j = 1, \ldots, 4$,

$$p\left(\sigma_j^2|\omega_{-\sigma_j^2}\right) \propto \left\{\prod_{i=1}^N p\left(b_{ij}|b_{ij-1}, \ldots, b_{i1}, s_i = s, \omega\right)\right\}.$$

- For $\Delta_{\alpha \cdot \theta \cdot \gamma}$,

$$p\left(\Delta_{\alpha \cdot \theta \cdot \gamma}|\omega_{-\Delta_{\alpha \cdot \theta \cdot \gamma}}\right) \propto \left\{\prod_{i=1}^N p\left(b_{ij}|b_{ij-1}, \ldots, b_{i1}, s_i = s, \omega\right)\right\}.$$

- For Δ_σ,

$$p\left(\Delta_\sigma\right) \propto \left\{\prod_{i=1}^N p\left(b_{ij}|b_{ij-1}, \ldots, b_{i1}, s_i = s, \omega\right)\right\}.$$

Since all full conditionals are intractable analytically except α, θ and γ, and not easily generated from, we have to construct suitable proposals for a Metropolis-Hastings step (Hastings, 1970; Gamerman, 1997). In practice, Gibbs sampling is implemented using WinBUGS (http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml). The MCMC algorithm simulates direct draws from the above full conditionals iteratively until convergence is achieved. Two long chains are used for the proposed model with different initial values.

Bayesian modeling has two model selection criteria, posterior predictive loss (PPL) (Gelfand and Ghosh, 1998) and deviance information criterion (DIC) (Spiegelhalter *et al.*, 2002). Both take into account goodness of fit while penalizing models for overfitting (a complexity penalty) (Daniels and Hogan, 2008). The posterior predictive loss is the average loss measured between the observed data and data predicted from the posterior predictive distribution. The comparison is based on a user-chosen loss function such as squared error loss. We use DIC for model comparison in this paper since it is not easy to choose a proper loss function in categorical data.

The DIC is a model-based criterion composed of a goodness of fit term and a penalty term that is computed as the mean deviance minus the deviance evaluated at the posterior mean. The DIC is defined as

$$\begin{aligned} \text{DIC} &= \overline{\text{Dev}(\omega)} + p_D \\ &= D(\tilde{\omega}) + p_D \\ &= 2\overline{\text{Dev}(\omega)} - D(\tilde{\omega}), \end{aligned}$$

where

$$p_D = \overline{\text{Dev}(\omega)} - D(\tilde{\omega})$$

and

$$\text{Dev}(\omega) = -2\log L(\omega|y).$$

The first term measures goodness of fit, $\overline{\text{Dev}(\omega)}$ is evaluated at the posterior mean of the parameter and the second term is the complexity penalty. $L(\omega|y)$ is the likelihood of $y = (y_1, \ldots, y_N)^T$, $\overline{\text{Dev}(\omega)}$ is the posterior mean deviance, $D(\tilde{\omega}) = \text{Dev}\{E_\omega(\omega|y)\}$, and The value of p_D is the effective number of parameters.
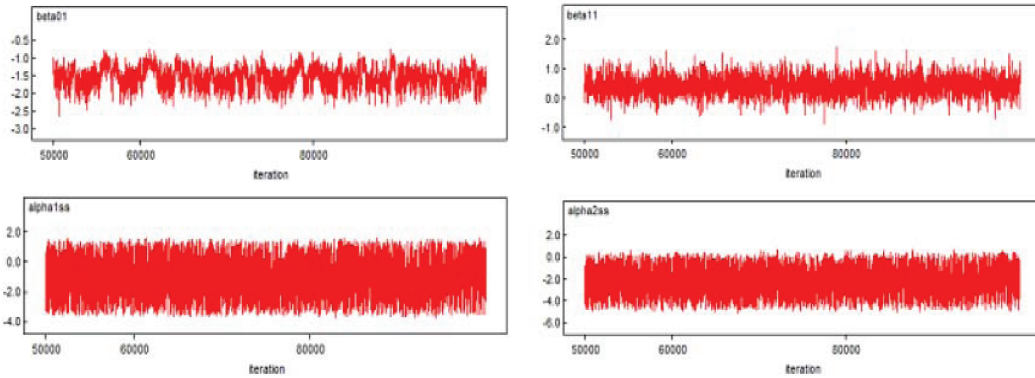
Figure 1: *Trace plots for $\beta_0^{(1)}$, $\beta_1^{(1)}$, $\alpha_1^{(1)}$, and $\alpha_3^{(1)}$ from WinBUGS.*

3.5. Model fits

We analyze pattern mixture models, composed of GLMMs within pattern, for repeated metabolic syndrome data of no treatment group. We using variable Y (1 if a participant has the metabolic syndrome and 0 otherwise) and Z (1 if a participant has smoking experience and 0 otherwise).

We fit two models for the distribution of the random effects b_i. Model 1 is given in (3.3)–(3.12). Models 2 reduces the parameters α, θ, and γ in Model 1; therefore, the distribution of b_i in Model 2 is given by

$$
\begin{aligned}
b_1|S &= 1 \sim N\left(0, \sigma_1^2\right), \\
b_1|S &= 2 \sim N\left(0, \sigma_1^2\right), \\
b_1|S &= 3 \sim N\left(0, \sigma_1^2\right), \\
b_1|S &= 4 \sim N\left(0, \sigma_1^2\right), \\
b_2|b_1, S &= 1 \sim N\left(\alpha_1^* b_1, \sigma_2^{*2}\right), \\
b_2|b_1, S &\geq 2 \sim N\left(\alpha_1^* b_1, \sigma_2^2\right), \\
b_3|b_1, b_2, S &= 1, 2 \sim N\left(\alpha_2^* b_1 + \theta_1^* b_2, \sigma_3^{*2}\right), \\
b_3|b_1, b_2, S &= 3, 4 \sim N\left(\alpha_2^* b_1 + \theta_1^* b_2, \sigma_3^2\right), \\
b_4|b_1, b_2, b_3, S &= 1, 2, 3 \sim N\left(\alpha_3^* b_1 + \theta_2^* b_2 + \gamma_1^* b_3, \sigma_4^{*2}\right), \\
b_4|b_1, b_2, b_3, S &= 4 \sim N\left(\alpha_3^* b_1 + \theta_2^* b_2 + \gamma_1^* b_3, \sigma_4^2\right).
\end{aligned}
$$

Each model is implemented using MCMC algorithm using WinBUGS 1.4.3. The posterior means were calculated with a sample size of 100,000 and burn-in period of 50,000. Figure 1 presents trace plots from WinBUGS and we observe that the lines of a chain mixes and crosses in trace plots; subsequently, convergence is ensured. The DIC values for two models, Models 1 and 2 are 10312.5 and 10314.8, respectively, and the two values are similar. By parsimonious rule, we take Model 1. Table 2 presents posterior means for parameters for Model 1. The bold-faced numbers are posterior means of $h(\xi_M, \Delta)$ which is given in (3.13)–(3.16).

Table 2: Metabolic syndrome study: posterior means for parameters from the proposed model

Parameter	Pattern (S)				Parameter	Pattern (S)			
	1	2	3	4		1	2	3	4
$\beta_0^{(S)}$	−1.5	−1.8	−2.1	−4.7	$\alpha_1^{(S)}$	**−1.0**	1.4	1.4	1.4
$\theta_2^{(S)}$	**−2.1**	**−2.1**	**−2.1**	0.3	$\alpha_2^{(S)}$	**−0.5**	**−0.5**	0.3	0.3
$\beta_1^{(S)}$	0.5	−0.1	−0.9	2.0	$\gamma_1^{(S)}$	**−1.9**	**−1.9**	**−1.9**	0.5
$\alpha_3^{(S)}$	**−2.3**	**−2.3**	**−2.3**	0.2	$\theta_1^{(S)}$	**−2.0**	**−2.0**	0.5	0.5

We consider respectively two models for smoker ($Z = 1$) and nonsmoker ($Z = 0$) given by

$$\text{logit}\left(P_{it}^{(S)}(b_{it})\right) = \beta_0^{(S)} + \beta_1^{(S)} + b_{it}, \tag{3.29}$$

$$\text{logit}\left(P_{it}^{(S)}(b_{it})\right) = \beta_0^{(S)} + b_{it}. \tag{3.30}$$

The differences between (3.20) and (3.21) is $\beta_1^{(S)}$. For patterns 1 and 4, the log odds of metabolic syndrome for smoker are 1.65 and 7.39 higher than that for non-smoker, respectively. However, For patterns 2 and 3, the log odds of metabolic syndrome is 0.9, 0.41 lower than that for non-smoker, respectively. The marginal estimates over patterns can be derived as a weighted average across pattern specific estimates. The estimated values for β_0 and β_1 are −3.51 and 1.20 respectively. It indicates that the log odds of metabolic syndrome for smokers are different from non-smokers.

4. Summary and Discussion

We propose pattern mixture models for longitudinal binary data with missingness using transition models or GLMMs. In the transition models, the serial dependence is explained via Markovian structure and the parameters are different over pattern. In the GLMMs, the effect of covariates on response is a specified conditional on random effects. The random effects covariance matrix for the random effects is different over patterns. All parameters are classified in both models into identified and non-identified parameters. The nonidentified parameters are reparameterized using sensitivity parameters and are assessed via sensitivity analysis.

The analysis of metabolic syndrome data indicated that most of participants dropped out after the second visit. We fit two presented models using Bayesian methodology. Both identifiable and non-identifiable parameters are estimated and sensitivity analysis uses series of point mass priors, starting with MAR and the number of sensitivity parameters are reduced based on assumption. The log odds of metabolic syndrome for smokers are different from that for non-smokers over patterns.

Acknowledgement

References

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association*, **88**, 125–134.

Daniels, M. and Hogan, J. (2008). Missing data in longitudinal studies: Strategies for Bayesian modeling and sensitivity analysis, *TEST*, **18**, 51–58.

Demirtas, H. and Schafer, J. L. (2003). On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out, *Statistics in Medicine*, **22**, 1113–1121.

Diggle, J. and Kenward, M. (1994). Informative dropout in longitudinal data analysis, *Journal of the Royal Statistical Society*, **43**, 49–93.

Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman & Hall, London.

Gelfand, A. and Ghosh, S. (1998). Model choice: A minimum posterior predictive loss approach, *Biometrika*, **53**, 1527–1537.

Gelman, A., Carlin, J., Stern, H. and Rubin, D. B. (2004). *Bayesian Data Analysis*, 2nd Edition, Chapman & Hall, London, UK, 1995.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*, **57**, 97–109.

Heckman, J. J. (1974). Shadow wages, market wages and labor supply, *Econometrica*, **42**, 679–693.

Hogan, J. and Laird, N. (1997). Model-based approaches to analysing incomplete longitudinal and failure time data, *Statistics in Medicine*, **16**, 259–272.

Kenward, M. and Molenberghs, G. (1999). Parametric models for incomplete continuous and categorical longitudinal data, *Statistical Methods in Medical Research*, **8**, 51.

Kim, J., Kim, E., Yi, H., Joo, S., Shin, K., Kim, J., Kimm, K. and Shin, C. (2006). Short-term incidence rate of hypertension in Korea middle-aged adults, *Journal of Hypertension*, **24**, 2177–2182.

Lee, K., Daniels, M. and Joo, Y. (2013). Flexible marginalized models for bivariate longitudinal ordinal data, *Biostatistics*, **14**, 462–476.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies, *Journal of the American Statistical Association*, **90**, 1112–1121.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd edition, John Wiley, New York.

Molenberghs, G. and Kenward, M. (2007). *Missing Data in Clinical Studies*, Wiley.

Rubin, D. B. (1976). Inference and missing data, *Biometrika*, **63**, 581–590.

Rubin, D. B. (1977). Formalizing subjective notions about the effect of nonrespondents in sample surveys, *Journal of the American Statistical Association*, **72**, 538–543.

Scharfstein, D. O., Rotnitzky, A. and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models, *Journal of the American Statistical Association*, **94**, 448.

Spiegelhalter, D. J., 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, **64**, 583-640, 651–653, 655, 658, 661, 671.