Title	Bayesian Modeling and Inference for Nonignorably Missing Longitudinal Binary Response Data with Applications to HIV Prevention Trials
Manuscript ID	SS-2016-0319.R1
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0319
Complete List of Authors	Joseph Ibrahim Jing Wu Ming-Hui Chen Elizabeth Schifano and Jeffrey Fisher
Corresponding Author	Joseph Ibrahim
E-mail	ibrahim@bios.unc.edu

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

Bayesian Modeling and Inference for Nonignorably Missing Longitudinal Binary Response Data with Applications to HIV Prevention Trials

Jing Wu*, Joseph G. Ibrahim‡, Ming-Hui Chen*,
Elizabeth D. Schifano*, and Jeffrey D. Fisher†

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

‡Department of Biostatistics, University of North Carolina,
McGavran-Greenberg Hall, Chapel Hill, NC, USA

† Department of Psychology and Institute for Collaboration on Health, Intervention, and Policy (InCHIP), University of Connecticut, Storrs, CT, USA

April 28, 2017

Abstract

Missing data are frequently encountered in longitudinal clinical trials. To better monitor and understand the progress over time, we must handle the missing data appropriately and thus examine whether the missing data mechanism is ignorable or nonignorable. In this article, we develop a new probit model for longitudinal binary response data. It resolves a challenging issue for estimating the variance of the random effects, and substantially improves the convergence and mixing of the Gibbs sampling algorithm. We further show that when improper uniform priors are specified for the regression coefficients of the joint multinomial model via a sequence of one-dimensional conditional distributions for the missing data indicators under nonignorable missingness, the joint posterior distribution is improper. A variation of Jeffreys prior is thus established as a remedy for the improper posterior distribution. In addition, an efficient Gibbs sampling algorithm is developed using a collapsing technique. Two model assessment criteria, the deviance information criterion (DIC) and the logarithm of the pseudomarginal likelihood (LPML), are used to guide the choices of prior specifications and to compare the models under different missing data mechanisms. An extensive simulation is conducted to investigate the empirical performance of the proposed methods. The proposed methodology is further illustrated using real data from an HIV prevention clinical trial.

Keywords: Probit Model; Latent Variable; Jeffreys Prior; Collapsed Gibbs Sampler; Identifiability; DIC; LPML.

1 Introduction

Intermittent missingness and dropout are frequently encountered in longitudinal studies. Intermittent missingness occurs when the subject returns to the study after missing one or more visits and dropout refers to the situation where the subject permanently withdraws from the study.

Little and Rubin (2002) classified the type of missingness into three categories, "Missing Completely at Random " (MCAR) is where the probability of missingness does not depend on either the observed or unobserved data. "Missing at Random" (MAR) is the situation where the probability of missingness does not depend on the unobserved data conditional on the observed data. "Missing Not at Random" (MNAR) is the setting in which the probability of missingness depends on the unobserved data. MCAR and MAR are typically referred to as ignorable missing data mechanisms since the missing data mechanism does not need to be included in the likelihood specification, while MNAR is referred to as a nonignorable missing mechanism for obtaining the maximum likelihood estimates. Nonignorable missing data is most frequently encountered in longitudinal studies, where data is gathered for the same subject repeatedly over time.

One approach for handling missing data is listwise deletion, in which all cases with missing values are deleted. This approach, however, introduces bias if the missingness is not MCAR. For MAR, inferential methods include maximum likelihood (Rubin, 1976; Ibrahim *et al.*, 1999; Newman, 2003; Ibrahim *et al.*, 2005), multiple imputation (Rubin, 2004; Royston and others, 2004; Sterne *et al.*, 2009) and weighted estimating equations (Robins and Rotnitzky, 1995; Preisser *et al.*, 2002). If the data are MNAR, one approach is to specify a parametric model for the missing data mechanism, and then jointly model the response variables and the missing data mechanism by incorporating them into the complete data log-likelihood. Three commonly used joint models are selection (Glynn *et al.*, 1986), pattern-mixture (Little, 1993), and shared-parameter models (Follmann and Wu, 1995).

Ibrahim *et al.* (2001) proposed a general joint multinomial model for the missing data mechanism for longitudinal data, which nicely accommodates nonignorable missing response data with nonmonotone missingness patterns. They also devised a Monte Carlo EM algorithm, and derived the analytical form of the E- and M-steps for the normal random effects model. Huang *et al.* (2005) provided theoretical justifications of model identifiability for generalized linear models with nonignorably missing covariates where they mainly focused on missing covariates rather than missing response measurements. Albert (2000) considered the transition model, which is appropriate if one is interested in how the response and covariates are related to the missingness path of each subject. He examined the setting of intermittent missingness and proposed a transition model for longitudinal binary data which allows for nonignorable intermittent missingness and dropout of each subject. However, the model does not allow for correlations between the response variable within each subject, and it also does not consider the fact that an intermittent missing value at time t must be followed by an observed value at some time point greater than t (otherwise, it would be a dropout).

One challenge of the probit mixed-effects regression model for longitudinal binary response data is the estimation of the variances of the random effects. In this paper, we propose a new reparameterization technique to develop a new probit model with latent variables. Our proposed model not only makes the variance for the random effects more identifiable but it also improves convergence and mixing of the Gibbs sampling algorithm, particularly for the parameters involved in the covariance matrix of the random effects. Following Ibrahim *et al.* (2001, 2005), we adopt a sequence of one-dimensional conditional distributions for the missing data indicators via a logistic regression model, and further show that the posterior distribution is improper if improper uniform priors are specified for the regression coefficients corresponding to the missing binary responses in the logistic regression models. To overcome this non-identifiability issue, we first specify normal priors for these regression coefficients and then use the DIC and LPML criteria to guide the choice of "optimal" normal priors for the regression coefficients. We further propose a variation of Jeffreys prior, which circumvents the identifiability issue all together. The proposed Jeffreys prior is attractive since it is relatively noninformative, guarantees that the joint posterior distribution is proper, and has similar performance as the "optimal" normal priors. Finally, the proposed joint model for the longitudinal binary responses and the missing data mechanism (ignorable or nonignorable) is computationally attractive since it allows us to conveniently sample missing binary responses and to apply the collapsed Gibbs technique (Liu, 1994) within the Gibbs sampling framework.

The remainder of this article is organized as follows. A brief description of the HIV prevention behavioral data is presented in Section 2. Section 3 introduces a new probit model with latent variables, and presents a joint multinomial model for the missing data indicators. In Section 4, we investigate and characterize the conditions for propriety of the joint posterior distribution, followed by a variation of Jeffreys prior as a remedy for impropriety of the posterior. In addition, we develop an efficient Gibbs sampling algorithm, and in the same section, provide a detailed formulation of the partial DIC and conditional LPML criteria in the presence of missing data. An extensive simulation is carried out in Section 5. In Section 6, we carry out a detailed analysis of the HIV prevention behavioral data. We conclude the paper with a brief discussion in Section 7.

2 HIV Prevention Behavioral Data

We consider data from an HIV prevention behavioral intervention clinical trial (Fisher *et al.*, 2014) in South Africa, where people living with HIV (PLWH) on antiretroviral therapy (ART) constitute a large population. The goal of this trial was to understand if a brief counseling intervention can significantly reduce HIV risk behavior among HIV-infected South Africans on ART. The data were collected from sixteen urban, peri-urban, and rural primary healthcare clinics and community health centers in the uMgungundlovu and uMkhanyakude health districts of KwaZulu-Natal, South Africa from June 2008 to May 2010. The sixteen health districts were then randomized to intervention (8 clinics) and standard of care (8 clinics) arms. The total number of HIV-infected participants on ART was 1891 (967 for intervention and 924 for standard of care).

Table 1: Characteristics of Study Participants (N=1875)

Characteristics (N=1875)	**Standard of Care (N=915)**	**Intervention (N=960)**	P
Lives in city or township			0.008
Yes	148 (16.17%)	202 (21.04%)	
No	767 (83.83%)	758 (78.96%)	
Cohabitates with sex partner			0.034
Yes	470 (51.37%)	445 (46.35%)	
No	445 (48.63%)	515 (53.65%)	
Meets with a counselor at clinic every 3 months or less			0.017
Yes	768 (83.93%)	764 (79.58%)	
No	147 (16.07%)	196 (20.42%)	
Reported drinking alcohol weekly or more frequently			<0.001
Yes	47 (5.14%)	16 (1.67%)	
No	868 (94.97%)	944 (98.33%)	
Depressed (modified CESD 11 score of 9 or more)			0.036
Yes	480 (52.46%)	551 (57.40%)	
No	435 (47.54%)	409 (42.60%)	
Gender			0.924
Female	511 (55.85%)	533 (55.52%)	
Male	404 (44.15%)	427 (44.48%)	
Median Age (IQR)	36 (31, 42)	36 (31, 43)	0.447

The final column indicates the p-values from the Mantel-Haenszel Chi-squared test(categorical covariates) and the Wilcoxon rank sum test (continuous covariates) for equality of proportions.

PLWH were invited to take part in the study and provided informed consent. Participation consisted of (1) completing audio computer- assisted self-interviews (ACASI) and interviewer-administered questionnaires at baseline, 6, 12, and 18 months, (2) providing biological samples assessing sexually transmitted infections (STIs) at baseline, 12, and 18 months, and (3) consenting to medical chart reviews for CD4 count, HIV viral load, STIs, and health status. As part of routine clinical care, participants in the intervention (n = 967) and standard of care (n = 924) arms received counseling from lay counselors concerning issues relevant to PLWH on ART (e.g., adherence education and counseling). Participants at the 8 intervention clinics (n = 967) received brief, theory and evidence-based, tailored, one-on-one counseling sessions with trained lay counselors concerning sexual risk behavior reduction. Standard of care participants received standard of care safer sex promotion messages from counselors, typically involving standard condom promotion messaging. Assessments were carried out by a different individual in a separate research setting at the 4 specified time points within the 18-month study.

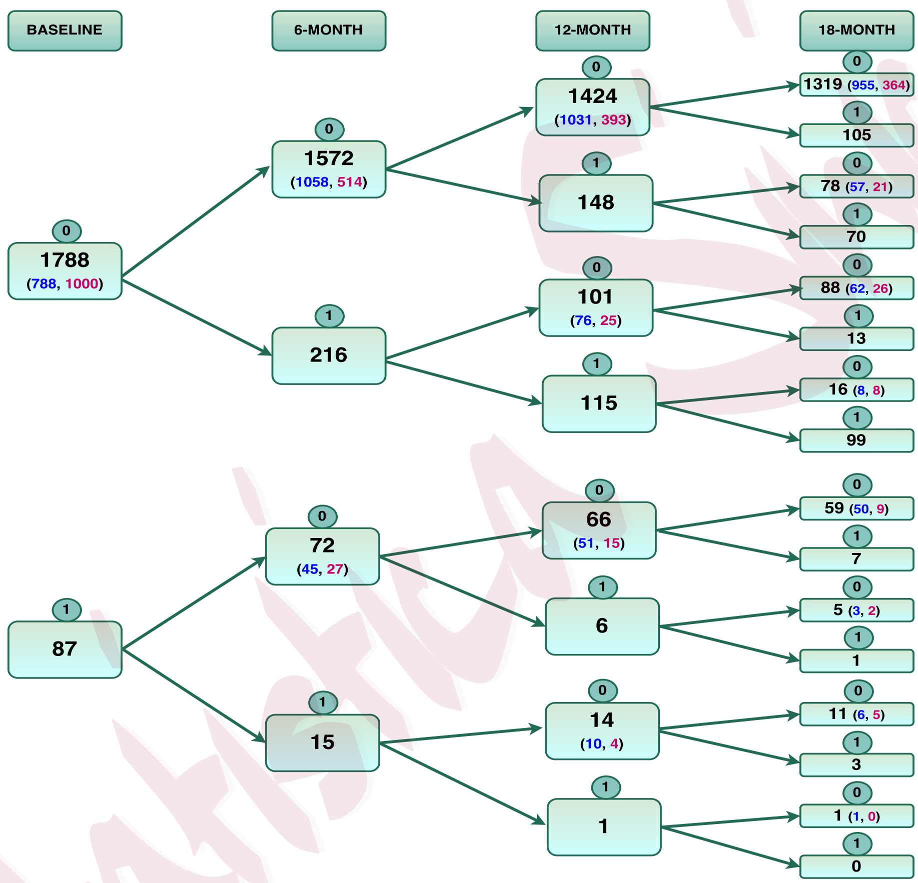

Figure 1: Path Diagram of the binary responses (any unprotected sex acts), where 0 in circle indicates observed and 1 in circle indicates missing; and the two numbers in parentheses indicate the number of zero counts (the first, blue) and the number of ones (the second, red) of the binary response variable at each visit on the specific path.

The longitudinal binary response variable is any ACASI-reported unprotected penile-vaginal or penile-anal sex acts in the past 4 weeks with partners of any HIV status, where 1 denotes the occurrence and 0 indicates otherwise. We excluded subjects who had missing values for the entire study, including baseline measurements from our analysis. We also excluded four subjects who had missing baseline covariates, so that the resulting number of subjects in our study cohort is 1875. Table 1 shows the characteristics of these 1875 PLWH, and Figure 1 visually presents the path diagram of the longitudinal binary response data (any unprotected sex acts). Determining whether missing responses are ignorable or nonignorable is of great practical interest in HIV intervention clinical trials, which greatly motivates our proposed methodology.

3 The Proposed Models

Suppose there are a total of T visits and K health districts in a clinical trial. Let y_t denote the measurement for a patient at visit t in the k^{th} health district ($1 \leq k \leq K$), and $\mathbf{y}_t = (y_0, y_1, \ldots, y_t)'$ denote the vector containing all the measurements up to and including visit t, for $t = 0, \ldots, T$, where y_0 represents the baseline measurement. Also, denote by z the intervention indicator such that $z = 0$ if the subject belongs to the control arm and $z = 1$ if the subject belongs to the intervention arm.

3.1 The Model for Longitudinal Binary Measurements

According to Verbeke (2005), for longitudinal measurements, it is often assumed that y_t follows a pre-specified distribution $F(\boldsymbol{\beta}, \epsilon_t)$, depending on covariates and is parameterized through a vector $\boldsymbol{\beta}$, common to all subjects, and subject-specific random effects ϵ_t. When y_t is binary, the probit mixed-effects regression model is assumed and given by

$$P(y_t = 1 | z, \mathbf{x}_1, k, \boldsymbol{\beta}^*, \tau^*, \zeta_k, \epsilon_t^*) = \Phi(z\beta_{1t}^* + \mathbf{x}_1' \boldsymbol{\beta}_{2t}^* + \tau^* \zeta_k + \epsilon_t^*), \tag{3.1}$$

for $t = 0, \ldots, T$, where Φ is the $N(0,1)$ cumulative distribution function, $\mathbf{x}_1$ is a vector of baseline covariates, $\boldsymbol{\beta}^* = (\beta_{1t}^*, \boldsymbol{\beta}_{2t}^{*'})'$ with β_{1t}^* denoting the regression coefficient corresponding to treatment condition and $\boldsymbol{\beta}_{2t}^*$ is the vector of regression coefficients corresponding to $\mathbf{x}_1$. Due to the design of the HIV prevention behavioral data that sixteen health districts were randomized instead of patients, we introduce random effects $\zeta_k \overset{i.i.d.}{\sim} N(0,1)$ with $\tau^{*2} (\tau^* > 0)$ being the variance, representing the random effect for all the patients from the k^{th} heath district, $k = 1, \ldots, K$. We further assume that $\boldsymbol{\epsilon}^* = (\epsilon_0^*, \epsilon_1^*, \ldots, \epsilon_T^*)' \sim N(\mathbf{0}, \sigma^2 \Sigma)$, where Σ is a $(T+1) \times (T+1)$ correlation matrix with $(s,t)^{th}$ entry $\rho^{|t-s|}$. However, under this formulation, the variance σ^2 of the random effects cannot be estimated.

To better see this identifiability problem, we obtain an equivalent representation of the model given in (3.1) by introducing the latent variables $\mathbf{w}^* = (w_0^*, \ldots, w_T^*)$. Following

Albert and Chib (1993), (3.1) can be reformulated as

$$y_t = \begin{cases} 1 & \text{if } w_t^* \geq 0, \\ 0 & \text{if } w_t^* < 0, \end{cases} \tag{3.2}$$

and

$$w_t^* \mid \epsilon_t^* \sim N(z\boldsymbol{\beta}_{1t}^* + \mathbf{x}_1'\boldsymbol{\beta}_{2t}^* + \tau^*\zeta_k + \epsilon_t^*, 1) \tag{3.3}$$

for $t = 0, 1, \ldots, T$, where $\boldsymbol{\epsilon}^* = (\epsilon_0^*, \epsilon_1^*, \ldots, \epsilon_T^*)' \sim N(\mathbf{0}, \sigma^2\Sigma)$.

First we note that y_t modeled in (3.2) is invariant with respect to the scale parameter (variance) of w_t^*. To be more specific, if we replace w_t^* in (3.3) by $C \cdot w_t^*$, where C is any nonnegative constant, (3.2) is still identical to (3.1). Therefore, the marginal variance of w_t^* as well as the marginal variance of $\boldsymbol{\epsilon}_t^*$ are not identifiable. Another issue with this model is that the marginal variance of each individual w_t^* given health districts, which is $1 + \sigma^2$, is partially confounded with the scale parameter σ^2 in the binary response model (See Kim *et al.* (2008) for a related discussion and Remark 3.1). These issues ultimately imply that $\boldsymbol{\beta}^*$ is essentially not identifiable and this leads to poor convergence of the Gibbs sampling algorithm. To circumvent these problems, we consider the following reparameterization:

$$w_t = \frac{w_t^*}{\sqrt{1+\sigma^2}}, \quad \boldsymbol{\beta}_t = \frac{\boldsymbol{\beta}_t^*}{\sqrt{1+\sigma^2}}, \quad \tau = \frac{\tau^*}{\sqrt{1+\sigma^2}}, \quad \epsilon_t = \frac{\epsilon_t^*}{\sqrt{1+\sigma^2}}. \tag{3.4}$$

After this reparameterization, we propose our equivalent but identifiable model as

$$P(y_t = 1 | z, \mathbf{x}_1, k, \boldsymbol{\beta}, \tau, \zeta_k, \epsilon_t) = \Phi((z\beta_{1t} + \mathbf{x}_1'\boldsymbol{\beta}_{2t} + \tau\zeta_k + \epsilon_t)\sqrt{1+\sigma^2}) = \pi_t, \tag{3.5}$$

or

$$y_t = \begin{cases} 1 & \text{if } w_t \geq 0, \\ 0 & \text{if } w_t < 0, \end{cases} \tag{3.6}$$

and

$$w_t \mid \epsilon_t \sim N(z\beta_{1t} + \mathbf{x}_1'\boldsymbol{\beta}_{2t} + \tau\zeta_k + \epsilon_t, \frac{1}{1+\sigma^2}) \tag{3.7}$$

for $t = 0, 1, \ldots, T$, where $\boldsymbol{\epsilon} = (\epsilon_0, \ldots, \epsilon_T)' \sim N(\mathbf{0}, \frac{\sigma^2}{1+\sigma^2}\Sigma)$. Under this new model, the marginal variance of $\mathbf{w}_t$ equals 1, leading to a better separation between $\boldsymbol{\beta}$ and σ^2, and improving convergence and mixing of the Gibbs sampling algorithm. For simplicity, we let α denote $\frac{\sigma^2}{1+\sigma^2}$ throughout the remainder of the paper.

The proposed model is attractive since (i) ϵ_t captures the dependence of the longitudinal measures, y_t, over time; (ii) the time-varying vector of coefficients $\boldsymbol{\beta}_t$ allows us to assess effectiveness of the intervention over time; (iii) the random effect $\boldsymbol{\zeta}$ adjusts for the effects of 16 health districts; and most importantly (iv) all the parameters involved in the model given by (3.5) or the model defined by (3.6) and (3.7) are identifiable.

Remark 3.1: After the reparameterization in (3.4), $\boldsymbol{\beta}_t$, as the ratio of $\boldsymbol{\beta}_t^*$ and $\sqrt{1+\sigma^2}$ is now identifiable. This implies that, in the original formulation of (3.3), a large value of σ^2 corresponds to large absolute values of the elements in $\boldsymbol{\beta}^*$ due to the dual role σ^2 plays in both the binary response and the latent variable model. It thus becomes difficult to interpret the meaning of $\boldsymbol{\beta}^*$, and leads to poor convergence of the Gibbs sampling algorithm. This phenomenon is also empirically observed in our analysis of the HIV data discussed in Section 2 by fitting the model defined by (3.2) and (3.3) without reparameterization, which further confirms the necessity of the reparameterization technique.

3.2 Missing Data Mechanism

Let $\mathbf{R}_T = (R_0, \ldots, R_T)'$ denote the vector of the missing data indicators. The missing data indicator, R_t, at time t is defined as

$$R_t = \begin{cases} 0 & \text{if } y_t \text{ is observed,} \\ 1 & \text{if } y_t \text{ is missing.} \end{cases}$$

Denoting $P(R_t = 1|\mathbf{R}_{t-1}, \mathbf{y}_t, z, \mathbf{x}_2, \boldsymbol{\gamma}_t) \triangleq P_t$, a logistic regression model is assumed for P_t:

$$\text{logit}(P_t) = \log\left(\frac{P_t}{1-P_t}\right) = z\gamma_{1t} + \mathbf{x}_2'\boldsymbol{\gamma}_{2t} + g(\mathbf{R}_{t-1}, \boldsymbol{\gamma}_{3t}) + h(\mathbf{y}_t, \boldsymbol{\gamma}_{4t}), \tag{3.8}$$

where $\mathbf{x}_2$ is a vector of baseline covariates, which may be different from $\mathbf{x}_1$, while g and h are certain linear functions. We set $g = 0$ when $t = 0$ since there are no previous missing indicators ($\mathbf{R}_{t-1}$). Following Ibrahim *et al.* (1999, 2005), we construct the joint distribution of $\mathbf{R}$ via a sequence of one-dimensional conditional distributions,

$$P(R_0 = r_0, \ldots, R_t = r_t|\mathbf{y}_t, z, \mathbf{x}_2, \boldsymbol{\gamma}) = \prod_{t=0}^{T} P_t^{\mathbf{1}(r_t=1)}(1-P_t)^{\mathbf{1}(r_t=0)}. \tag{3.9}$$

Remark 3.2: If we assume that $P(R_t = m|R_{t-1} = l, \mathbf{y}_t, z, \mathbf{x}_2, \boldsymbol{\gamma}_t)$ depends on the longitudinal measures only through the current and previous visits, we simply take $h(\mathbf{y}_t, \boldsymbol{\gamma}_{4t}) = \gamma_{4t1}y_{t-1} + \gamma_{4t2}y_t$ in (3.8). The model in (3.9) implies nonignorable missingness due to the existence of intermittent missingness and dropout. We may also let $h(\mathbf{y}_t, \boldsymbol{\gamma}_{4t}) = 0$ if the missingness is ignorable. (See Section 6 for further discussion.)

Remark 3.3: For $t > 0$, we may choose $g(\mathbf{R}_{t-1}, \boldsymbol{\gamma}_{3t}) = \mathbf{R}_{t-1}'\boldsymbol{\gamma}_{3t}$, which depends on all of the previous missingness indicators. In this paper, we set $g(\mathbf{R}_{t-1}, \boldsymbol{\gamma}_{3t}) = \sum_{j=0}^{t-1} R_j\boldsymbol{\gamma}_{3t}$. The new covariate $\sum_{j=0}^{t-1} R_j$ captures the cumulative number of missing response indicators, reduces the number of nuisance parameters for modeling the missing data mechanism, and makes the nonignorable missing data mechanism more identifiable (See Section 4.2).

4 Bayesian Inference

4.1 The Likelihood Function

Suppose there are n subjects and assume that $(z_i, k_i, \mathbf{x}_{1i}, \mathbf{x}_{2i})$ are completely observed, for all $i = 1, \ldots, n$. Let $\mathbf{y}_{\text{obs}} = (\mathbf{y}'_{1,\text{obs}}, \ldots, \mathbf{y}'_{n,\text{obs}})'$ and $\mathbf{y}_{\text{mis}} = (\mathbf{y}'_{1,\text{mis}}, \ldots, \mathbf{y}'_{n,\text{mis}})'$, where $(\mathbf{y}_{i,\text{obs}}, \mathbf{y}_{i,\text{mis}})$ are the observed and missing binary responses for the i^{th} subject.

Let $\mathbf{y}_i = (y_{i0}, \ldots, y_{iT})$, and $\mathbf{R}_{iT}$ denote the collection of all missing data indicators $\mathbf{R}_{iT} = (R_{i0}, \ldots, R_{iT})$. Denote by $D_c = \{\mathbf{y}_i, z_i, k_i, \mathbf{x}_{1i}, \mathbf{x}_{2i}, \zeta_{k_i}, \boldsymbol{\epsilon}_i, \mathbf{w}_i, \mathbf{R}_i, i = 1, \ldots, n\}$ the set of complete data and $D_{\text{obs}} = \{\mathbf{y}_{i,\text{obs}}, z_i, k_i, \mathbf{x}_{1i},\ \mathbf{x}_{2i}, \mathbf{R}_i, i = 1, \ldots, n\}$ is the set of observed data. Denote by f_y and $f_{\mathbf{R}}$ the marginal densities of $\mathbf{y}$ and $\mathbf{R}$, respectively. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \alpha, \tau, \rho)$ denote the collection of all model parameters.

Let $[A|B]$ denote the conditional distribution of A given B. We model the observed data through the sequence of conditional distributions $[\mathbf{y}][\mathbf{R}|\mathbf{y}]$. The complete data likelihood function is therefore given by

$$\mathcal{L}(\boldsymbol{\theta}|D_c) = \prod_{i=1}^{n} \Big\{ f_y(\mathbf{y}_i|z_i, \mathbf{x}_{1i}, k_i, \zeta_{k_i}, \boldsymbol{\epsilon}_i, \mathbf{w}_i, \boldsymbol{\theta}) f_{\mathbf{R}|\mathbf{y}}(\mathbf{R}_{iT}|\mathbf{y}_i, z_i, \mathbf{x}_{2i}, \boldsymbol{\theta}) \Big\}$$
$$= \prod_{i=1}^{n} \Big\{ \prod_{t=0}^{T} \mathbf{1}(w_{it} \geq 0)^{y_{it}} \mathbf{1}(w_{it} < 0)^{1-y_{it}} \frac{1}{\sqrt{2\pi(1-\alpha)}} \exp\{-\frac{(w_{it} - z_i\boldsymbol{\beta}_{1t} - \mathbf{x}'_{1i}\boldsymbol{\beta}_{2t} - \tau\zeta_{k_i} - \epsilon_{it})^2}{2(1-\alpha)}\}$$
$$P_{it}{}^{\mathbf{1}(r_{it}=1)}(1 - P_{it})^{\mathbf{1}(r_{it}=0)} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\zeta_{k_i}^2}{2}\right) \Big\} \frac{1}{\sqrt{2\pi|\alpha\Sigma|}} \exp\Big\{ -\frac{1}{2\alpha}\boldsymbol{\epsilon}'_i\Sigma^{-1}\boldsymbol{\epsilon}_i\Big\}. \tag{4.1}$$

After integrating out the missing longitudinal responses $\mathbf{y}_{i,\text{mis}}$, ζ_{k_i}, $\boldsymbol{\epsilon}_i$, and the latent variables $\mathbf{w}_i$, the observed data likelihood function is given by

$$\mathcal{L}(\boldsymbol{\theta}|D_{\text{obs}}) = \sum_{\mathbf{y}_{\text{mis}}} \int \prod_{i=1}^{n} \Big\{ \prod_{t=0}^{T} \mathbf{1}(w_{it} \geq 0)^{y_{it}} \mathbf{1}(w_{it} < 0)^{1-y_{it}}$$
$$\frac{1}{\sqrt{2\pi(1-\alpha)}} \exp\{-\frac{(w_{it} - z_i\boldsymbol{\beta}_{1t} - \mathbf{x}'_{1i}\boldsymbol{\beta}_{2t} - \tau\zeta_{k_i} - \epsilon_{it})^2}{2(1-\alpha)}\} d\mathbf{w} P_{it}{}^{\mathbf{1}(r_{it}=1)}(1 - P_{it})^{\mathbf{1}(r_{it}=0)}$$
$$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\zeta_{k_i}^2}{2}\right) d\boldsymbol{\zeta} \Big\} \frac{1}{\sqrt{2\pi|\alpha\Sigma|}} \exp\Big\{ -\frac{1}{2\alpha}\boldsymbol{\epsilon}'_i\Sigma^{-1}\boldsymbol{\epsilon}_i\Big\} d\boldsymbol{\epsilon}. \tag{4.2}$$

4.2 Prior and Posterior Distributions

We assume that the joint prior density can be expressed as

$$\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\beta})\pi(\boldsymbol{\gamma})\pi(\alpha)\pi(\tau)\pi(\rho).$$

The joint posterior based on the observed data D_{obs} is written as

$$\pi(\boldsymbol{\theta}|D_{\text{obs}}) \propto \mathcal{L}(\boldsymbol{\theta}|D_{\text{obs}})\pi(\boldsymbol{\theta}). \tag{4.3}$$

We first establish a useful proposition regarding the propriety of the posterior distribution when an improper uniform prior is assumed for $\boldsymbol{\gamma}$.

Proposition 4.1 *Suppose we take $\pi(\boldsymbol{\gamma}) \propto 1$, the joint posterior in (4.3) is improper regardless of whether $\pi(\boldsymbol{\beta}, \alpha, \tau, \rho)$ is proper or improper.*

A sketch of the proof of the proposition is given in Appendix A. From Proposition 4.1, the joint posterior distribution is improper if $\pi(\boldsymbol{\gamma}) \propto 1$. The next proposition, based on Chen and Shao (2001), states that under some mild conditions, the joint posterior is proper if $\pi(\boldsymbol{\gamma})$ is proper, but $\pi(\boldsymbol{\beta}, \alpha, \tau, \rho) \propto 1$.

Let $\mathbf{Z}_i$ be the $(T+1) \times (T+1)$ diagonal matrix with diagonal element being z_i, $\mathbf{X}_{1i}$ is the matrix with all the row vectors equal $\mathbf{x}'_{1i}$, and $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_T)'$ is a vector of length p. Denote by $I_c = \{i|R_{i0} = 0, \ldots, R_{iT} = 0\}$ the set of observations with no missing visits, and $\tilde{i} = (i-1)(T+1) + (t+1)$, for $1 \le i \le n$, $0 \le t \le T$. Let $\boldsymbol{\epsilon} = (\epsilon'_i, i \in I_c)'$, $\mathbf{u}_i = (u_{i0}, \ldots, u_{i,T})'$, $\mathbf{u} = (\mathbf{u}'_i, i \in I_c)'$, where the u_{it}'s are i.i.d $N(0,1)$ random variables. Let $\mathbf{X}^* = \{(\mathbf{Z}_i, \mathbf{X}_{1i})', i \in I_c\}'$ be the design matrix, where each row vector is defined as $\mathbf{x}'_{\tilde{i}}$. We further introduce $\mathbf{X}^*_{\text{obs}}$ to be the matrix with rows equal $(1 - y_{it})x'_{\tilde{i}}$, such that $i \in I_c$.

Proposition 4.2 *Suppose we take $\pi(\boldsymbol{\gamma})$ to be a proper prior, let $\pi(\tau)$ be a proper prior with a finite p^{th} moment, and specify improper uniform priors for the other parameters. The joint posterior in (4.3) is proper if the following conditions are satisfied: (C1) $\mathbf{X}^*$ is of full rank; and (C2) there exists a positive vector a, i.e., each component $a_i > 0$, such that $\mathbf{X}^*_{\text{obs}}{}'a = 0$.*

Next, we consider Jeffreys prior (Jeffreys, 1946) regarding $\boldsymbol{\gamma}$. Due to the involvement of the missing data in the design matrix, the conventional Jeffreys prior is computationally infeasible. However, we observe that Jeffreys prior based on a certain subset of the data is not only computationally feasible, but also leads to a proper posterior distribution (Chen *et al.*, 2008). Thus, we propose a variation of Jeffreys prior, which is analytically attractive. To be specific, we select a certain observed subset, denoted by $\tilde{D}_{\text{obs}}$, such that the likelihood function of the parameters does not involve any missing data. The logarithm of the joint likelihood function in (4.2) based on $\tilde{D}_{\text{obs}}$ is given by

$$\begin{aligned}\ell(\boldsymbol{\theta}|\tilde{\mathbf{D}}_{\text{obs}}) = &\log \int \prod_{(\mathbf{i},\mathbf{t})\in\tilde{\mathbf{D}}_{\text{obs}}} \mathbf{1}(\mathbf{w_{it}} \geq \mathbf{0})^{\mathbf{y_{it}}} \mathbf{1}(\mathbf{w_{it}} < \mathbf{0})^{\mathbf{1-y_{it}}} \\ &\frac{1}{\sqrt{2\pi(1-\alpha)}} \exp\{-\frac{(w_{it} - z_i\boldsymbol{\beta}_{1t} - \mathbf{x}_{1i}'\boldsymbol{\beta}_{2t} - \tau\zeta_{k_i} - \epsilon_{it})^2}{2(1-\alpha)}\} d\mathbf{w} \\ &\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\zeta_{k_i}^2}{2}\right) d\boldsymbol{\zeta} \frac{1}{\sqrt{2\pi|\alpha\Sigma|}} \exp\{-\frac{1}{2\alpha}\boldsymbol{\epsilon}_i'\Sigma^{-1}\boldsymbol{\epsilon}_i\} d\boldsymbol{\epsilon} \\ &+ \log \prod_{(i,t)\in\tilde{D}_{\text{obs}}} {P_{it}}^{\mathbf{1}(r_{it}=1)}(1-P_{it})^{\mathbf{1}(r_{it}=0)}. \end{aligned} \tag{4.4}$$

For $\boldsymbol{\gamma}_t$ at visit t, we use a different observed subset to construct the prior, aiming to utilize as many observations as possible. Indeed, the idea of using a subset of the data is equivalent to selecting the corresponding terms from the log-likelihood function. That is, if we take $h(\mathbf{y}_t, \boldsymbol{\gamma}_{4t}) = \gamma_{4t}y_t$ for $t = 0$, and $h(\mathbf{y}_t, \boldsymbol{\gamma}_{4t}) = \gamma_{4t1}y_{t-1} + \gamma_{4t2}y_t$ for $t > 0$ in (3.8), the log-likelihood of $\boldsymbol{\gamma}_t$ based on this subset of the data is given by

$$\begin{aligned}\ell(\boldsymbol{\gamma_t}|\mathbf{D_c}) &= \begin{cases} \sum_{i=1}^n \log\left\{\left[{P_{it}}^{\mathbf{1}(r_{it}=1)}(1-P_{it})^{\mathbf{1}(r_{it}=0)}\right]^{\mathbf{1}(r_{it}=0)}\right\} & t=0, \\ \sum_{i=1}^n \log\left\{\left[{P_{it}}^{\mathbf{1}(r_{it}=1)}(1-P_{it})^{\mathbf{1}(r_{it}=0)}\right]^{\mathbf{1}(r_{it-1}=0)\mathbf{1}(r_{it}=0)}\right\} & t>0, \end{cases} \\ &= \begin{cases} \sum_{i=1}^n \mathbf{1}(r_{it}=0)\log(1-P_{it}) & t=0, \\ \sum_{i=1}^n \mathbf{1}(r_{it-1}=0)\mathbf{1}(r_{it}=0)\log(1-P_{it}) & t>0. \end{cases}\end{aligned}$$

We now specify the joint prior distribution for $\boldsymbol{\gamma}_t$ as

$$\pi(\boldsymbol{\gamma}_t) \propto |\mathbf{X}_t^{*\prime}\mathbf{D}_t\mathbf{X}_t^*|^{1/2}, \tag{4.5}$$

where

$$\mathbf{X}_t^* = \begin{cases} \left[\mathbf{1}(r_{it}=0)\mathbf{X}_{it}^* : i=1,\ldots,n\right]' & t=0, \\ \left[\mathbf{1}(r_{it-1}=0)\mathbf{1}(r_{it}=0)\mathbf{X}_{it}^* : i=1,\ldots,n\right]' & t>0, \end{cases}$$

$|.|$ represents the determinant of a matrix, $\mathbf{X}_{it}^* = (z, \mathbf{x}_2', \mathbf{y}_{it})'$ if $t = 0$, and $\mathbf{X}_{it}^* = (z, \mathbf{x}_2', \sum_{j=0}^{t-1} R_j, \mathbf{y}_{it-1}, \mathbf{y}_{it})'$ for $t > 1$. For $t = 1$, since $\sum_{j=0}^{t-1} R_j = R_0 = 0$ for the subjects within this subset, an improper uniform prior is essentially assumed for γ_{3t} in $\pi(\boldsymbol{\gamma}_t)$ defined by (4.5) while Jeffreys prior is constructed for the other parameters in $\boldsymbol{\gamma}_t$ such that $\mathbf{X}_{it}^* = (z, \mathbf{x}_2', \mathbf{y}_{it-1}, \mathbf{y}_{it})'$. Also, in (4.5), $\mathbf{D}_t$ is an $n \times n$ diagonal matrix with diagonal elements being $P_{it}(1 - P_{it})$. If the design matrix $\mathbf{X}_t^*$ is of full column rank (Chen *et al.*, 2008), the prior for the corresponding parameters in $\boldsymbol{\gamma}_t$ is proper. In addition, we specify improper uniform priors for $(\boldsymbol{\beta}, \alpha, \rho)$, and a truncated normal prior for τ.

4.3 Computational Development

The joint posterior distribution of $(\boldsymbol{\theta}, \mathbf{y}_{\text{mis}})$ based on the observed data is given by

$$\pi(\boldsymbol{\theta}, \mathbf{y}_{\text{mis}}|D_{\text{obs}}) \propto \mathcal{L}(\boldsymbol{\theta}|D_c)\pi(\boldsymbol{\theta}), \tag{4.6}$$

where $\mathcal{L}(\boldsymbol{\theta}|D_c)$ is defined in (4.1). Thus, the joint posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \alpha, \tau, \rho)$ is written as

$$\begin{aligned}
&\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \alpha, \rho, \tau, \mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, |D_{\text{obs}}) \\
&\propto \prod_{i=1}^{n}\prod_{t=0}^{T}\left\{\mathbf{1}(w_{it} \geq 0)^{y_{it}}\mathbf{1}(w_{it} < 0)^{1-y_{it}}P_{it}{}^{\mathbf{1}(r_{it}=1)}(1-P_{it})^{\mathbf{1}(r_{it}=0)}\right\} \\
&(1-\alpha)^{-n(T+1)/2}\prod_{i=1}^{n}\prod_{t=0}^{T}\exp\left\{-\frac{(w_{it} - z_i\boldsymbol{\beta}_{1t} - \mathbf{x}_{1i}'\boldsymbol{\beta}_{2t} - \tau\zeta_{k_i} - \epsilon_{it})^2}{2(1-\alpha)}\right\}\prod_{i=1}^{n}\prod_{t=0}^{T}\exp\left(-\frac{\zeta_{k_i}^2}{2}\right) \\
&(\alpha)^{-n(T+1)/2}\prod_{i}^{n}|\Sigma|^{-1/2}\exp\left\{-\frac{1}{2\alpha}\boldsymbol{\epsilon}_i'\Sigma^{-1}\boldsymbol{\epsilon}_i\right\}\pi(\boldsymbol{\beta})\pi(\boldsymbol{\gamma})\pi(\alpha)\pi(\tau)\pi(\rho).
\end{aligned} \tag{4.7}$$

The Gibbs sampling algorithm requires sampling from the following full conditional distributions in turn:

$$\begin{array}{llll}
\text{(i)} & [\mathbf{y}_{\text{mis}}, \boldsymbol{\gamma}|\mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \alpha, \tau, \rho, D_{\text{obs}}]; & \text{(ii)} & [\mathbf{w}, \boldsymbol{\beta}|\mathbf{y}_{\text{mis}}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \alpha, \tau, \rho, D_{\text{obs}}]; \\
\text{(iii)} & [\alpha, \rho|\mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \tau, D_{\text{obs}}]; & \text{(iv)} & [\boldsymbol{\epsilon}|\mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \alpha, \tau, \rho, D_{\text{obs}}]; \\
\text{(v)} & [\tau|\mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \alpha, \rho, D_{\text{obs}}]; & \text{(vi)} & [\boldsymbol{\zeta}|\mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\epsilon}, \alpha, \tau, \rho, D_{\text{obs}}].
\end{array} \tag{4.8}$$

For (i), we first collapse out the latent random variables $\mathbf{w}$ via the following identity:

$$\begin{aligned}
&[\mathbf{y}_{\text{mis}}, \boldsymbol{\gamma}, \mathbf{w}, \boldsymbol{\beta}|\boldsymbol{\zeta}, \boldsymbol{\epsilon}, \alpha, \tau, \rho, D_{\text{obs}}] = [\mathbf{y}_{\text{mis}}, \boldsymbol{\gamma}|\boldsymbol{\beta}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \alpha, \tau, \rho, D_{\text{obs}}][\mathbf{w}, \boldsymbol{\beta}|\mathbf{y}_{\text{mis}}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \alpha, \tau, \rho, D_{\text{obs}}] \\
&= [\mathbf{y}_{\text{mis}}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \alpha, \tau, \rho, D_{\text{obs}}][\boldsymbol{\gamma}|\mathbf{y}_{\text{mis}}, D_{\text{obs}}][\mathbf{w}, \boldsymbol{\beta}|\mathbf{y}_{\text{mis}}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \alpha, \tau, \rho, D_{\text{obs}}],
\end{aligned} \tag{4.9}$$

and then run a sub-Gibbs sampling algorithm to sample from the following full conditional distributions in turn: (ia)$[\mathbf{y}_{\text{mis}}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \alpha, \tau, \rho, D_{\text{obs}}]$ and (ib)$[\boldsymbol{\gamma}|\mathbf{y}_{\text{mis}}, D_{\text{obs}}]$.

Sampling $\mathbf{w}$ and $\boldsymbol{\beta}$ in (ii) are straightforward since the components of $\mathbf{w}$ are conditionally independent truncated normal random variables, and $\boldsymbol{\beta}$, conditional on the other parameters and variables, follows a multivariate normal distribution.

The posterior distribution of (α, ρ) in the binary response model is highly dependent on the random effects $\boldsymbol{\epsilon}$. Directly sampling (α, ρ) from their full conditional distributions will lead to slow convergence and poor mixing of the Gibbs sampling algorithm. Due to the introduction of the probit link and the latent variables $\mathbf{w}$, we are able to analytically integrate out $\boldsymbol{\epsilon}$. For (iii), we again apply the collapsed Gibbs technique through the identity:

$$[\alpha, \rho, \boldsymbol{\epsilon}|\mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \tau, D_{\text{obs}}] = [\alpha, \rho|\mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \tau, D_{\text{obs}}][\boldsymbol{\epsilon}|\mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \alpha, \tau, \rho, D_{\text{obs}}]. \quad (4.10)$$

Sampling $\boldsymbol{\epsilon}$ in (iv) is also straightforward since the $\boldsymbol{\epsilon}_t$ are independent multivariate normal random variables conditional on the other parameters and variables.

Below, we briefly explain how to sample from these full conditional distributions.

Step (ia). For each missing response $y_{it,\text{mis}}$, we compute q_{it} as

$$q_{it} = \Big\{ \pi_{it} \prod_{j=t}^{T_0} P(r_{ij}|\mathbf{r}_{ij-1}, \mathbf{y}_{ij}, y_{it} = 1, z, \mathbf{x}_2, \boldsymbol{\gamma}) + (1 - \pi_{it}) \prod_{j=t}^{T_0} P(r_{ij}|\mathbf{r}_{ij-1}, \mathbf{y}_{ij}, y_{it} = 0, z, \mathbf{x}_2, \boldsymbol{\gamma}) \Big\}^{-1} \pi_{it} \prod_{j=t}^{T_0} P(r_{ij}|\mathbf{r}_{ij-1}, \mathbf{y}_{ij}, y_{it} = 1, z, \mathbf{x}_2, \boldsymbol{\gamma}),$$

where $T_0 = min(t+1, T)$, it refers to the t^{th} visit for the i^{th} observation, π_{it} is introduced in (3.5), and $P(r_{ij}|\mathbf{r}_{ij-1}, \mathbf{y}_{ij}, z, \mathbf{x}_2, \boldsymbol{\gamma})$ is given in (3.8). We next sample y_{it} from a Bernoulli(q_{it}) distribution.

Step (ib). We write the full conditional distribution of $\boldsymbol{\gamma}$ as

$$\pi(\boldsymbol{\gamma}_t|\mathbf{y}_{\text{mis}}, D_{\text{obs}}) \propto \prod_{i=1}^{n} P_{it}^{\mathbf{1}(r_{it}=1)}(1 - P_{it})^{\mathbf{1}(r_{it}=0)}\pi(\boldsymbol{\gamma}_t),$$

where P_{it} is established in (3.8). Let $\pi(\boldsymbol{\gamma})$ be the Jeffreys prior constructed in Section 4.2. We cannot use adaptive rejection sampling since Jeffreys prior is not log-concave (Chen *et al.*, 2008). Thus, we use the localized Metropolis algorithm to sample $\boldsymbol{\gamma}$.

Step (iia). We simply draw w_{it} from a truncated $N(z_i\beta_{1t} + \mathbf{x}_{1i}'\boldsymbol{\beta}_{2t} + \tau\zeta_{k_i} + \epsilon_{it}, 1 - \alpha)$ distribution given y_{it}, for $i = 1, \ldots, n$, and $t = 0, \ldots, T$.

Step (iib). Let $\tilde{\mathbf{X}}_i = (z_i, \mathbf{x}_{1i}')'$. Assuming $\pi(\boldsymbol{\beta}_t) \propto 1$, we sample $\boldsymbol{\beta}_t|\mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \alpha, \tau, \rho, D_{\text{obs}}$ for $t = 0, \ldots, T$ from

$$N\Big(\big(\sum_{i=1}^{n} \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\big)^{-1} \sum_{i=1}^{n} \tilde{\mathbf{X}}_i'(w_{it} - \tau\zeta_{k_i} - \epsilon_{it}), \big(\sum_{i=1}^{n} \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\big)^{-1}(1 - \alpha)\Big).$$

Step (iii). Let $\mu_{1i} = (w_{i0} - z_i\beta_{10} - \mathbf{x}'_{1i}\boldsymbol{\beta}_{20} - \tau\zeta_{k_i}, \ldots, w_{iT} - z_i\beta_{1T} - \mathbf{x}'_{1i}\boldsymbol{\beta}_{2T} - \tau\zeta_{k_i})'$ and ${\Sigma_1}^{-1} = \frac{1}{\alpha}\Sigma^{-1} + \frac{1}{1-\alpha}\mathbf{I}$. The joint full conditional distribution $[\alpha, \rho|\mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \tau, D_{\text{obs}}]$ is given by

$$\begin{aligned}
&\pi(\alpha, \rho|\mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \tau, D_{\text{obs}}) \\
&\propto \{\alpha(1-\alpha)\}^{-\frac{n(T+1)}{2}}|\Sigma|^{-\frac{n}{2}}\pi(\alpha)\pi(\rho) \\
&\quad \prod_{i=1}^{n} \exp\Big\{-\frac{\boldsymbol{\epsilon}'_i(\frac{1}{\alpha}\Sigma^{-1} + \frac{1}{1-\alpha}\mathbf{I})\boldsymbol{\epsilon}_i - \frac{2}{1-\alpha}\mu'_{1i}\boldsymbol{\epsilon}_i + \frac{1}{1-\alpha}\mu'_{1i}\mu_{1i}}{2}\Big\} \\
&\propto \{\alpha(1-\alpha)\}^{-\frac{n(T+1)}{2}}|\Sigma|^{-\frac{n}{2}}\pi(\alpha)\pi(\rho)\prod_{i=1}^{n}\exp\Big(\frac{\frac{1}{(1-\alpha)^2}\mu'_{1i}\Sigma_1\mu_{1i} - \frac{1}{1-\alpha}\mu'_{1i}\mu_{1i}}{2}\Big) \\
&\quad \prod_{i=1}^{n}\exp\Big\{-\frac{(\boldsymbol{\epsilon}_i - \frac{1}{1-\alpha}\Sigma_1\mu_{1i})'{\Sigma_1}^{-1}(\boldsymbol{\epsilon}_i - \frac{1}{1-\alpha}\Sigma_1\mu_{1i})}{2}\Big\}.
\end{aligned}$$

We next integrate out $\boldsymbol{\epsilon}$, and the joint full conditional distribution simplifies to

$$\begin{aligned}
&\pi(\alpha, \rho|\mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \tau, D_{\text{obs}}) \\
&\propto \{\alpha(1-\alpha)\}^{-\frac{n(T+1)}{2}}|\Sigma|^{-\frac{n}{2}}|\Sigma_1|^{\frac{n}{2}}\prod_{i=1}^{n}\exp\Big(\frac{\frac{1}{(1-\alpha)^2}\mu'_{1i}\Sigma_1\mu_{1i} - \frac{1}{1-\alpha}\mu'_{1i}\mu_{1i}}{2}\Big)\pi(\alpha)\pi(\rho).
\end{aligned}$$

(a). The full conditional distribution of α is given by

$$\begin{aligned}
\pi(\alpha|\mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \tau, \rho, D_{\text{obs}}) &\propto \{\alpha(1-\alpha)\}^{-\frac{n(T+1)}{2}}|\Sigma_1|^{\frac{n}{2}} \\
&\quad \prod_{i=1}^{n}\exp\Big(\frac{\frac{1}{(1-\alpha)^2}\mu'_{1i}\Sigma_1\mu_{1i} - \frac{1}{1-\alpha}\mu'_{1i}\mu_{1i}}{2}\Big)\pi(\alpha).
\end{aligned}$$

Since α is always between 0 and 1 exclusively, we introduce δ such that

$$\alpha = \frac{1}{1+e^{-\delta}}$$

with support on $(-\infty, \infty)$ to indirectly sample α. Thus

$$\pi(\delta|\mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \tau, \rho, D_{\text{obs}}) = \pi(\alpha|\mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \tau, \rho, D_{\text{obs}})\frac{e^{\delta}}{(1+e^{\delta})^2}.$$

Under a uniform prior specified for α, we use the localized Metropolis algorithm to sample δ, and then convert it back to α.

(b). The full conditional distribution of ρ is given by

$$\pi(\rho|\mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \alpha, \tau, D_{\text{obs}}) \propto |\Sigma|^{-\frac{n}{2}}|\Sigma_1|^{\frac{n}{2}} \prod_{i=1}^{n} \exp(\frac{\frac{1}{(1-\alpha)^2}\mu_{1i}'\Sigma_1\mu_{1i}}{2})\pi(\rho).$$

Since $-1 < \rho < 1$, we use a "de-constraining" transformation to sample ρ (Chen *et al.*, 2000):

$$\rho = \frac{-1+e^{\xi}}{1+e^{\xi}} \quad -\infty < \xi < \infty.$$

Thus

$$\pi(\xi|\mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \alpha, \tau, D_{\text{obs}}) = \pi(\rho|\mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \alpha, \tau, D_{\text{obs}})\frac{2e^{\xi}}{(1+e^{\xi})^2}.$$

Assume that a Uniform$(-1, 1)$ prior is specified for ρ. Since $\pi(\xi|\boldsymbol{\epsilon}, \boldsymbol{\beta}, \alpha, \mathbf{y}_{\text{mis}}, D_{\text{obs}})$ is not log-concave, we again use the localized Metropolis algorithm to sample ξ, and then convert it back to ρ.

Step (iv). Based on the derivation in Step (iii), draw $\boldsymbol{\epsilon}_i$ from a $N\left(\frac{1}{1-\alpha}\Sigma_1\mu_{1i}, \Sigma_1\right)$.

Step (v). The full conditional distribution of τ is given by

$$\pi(\tau|\mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \alpha, \rho, D_{\text{obs}}) \propto \exp\left\{-\frac{\sum_{i=1}^{n}\sum_{t=0}^{T}(w_{it} - z_i\boldsymbol{\beta}_{1t} - \mathbf{x}_{1i}'\boldsymbol{\beta}_{2t} - \tau\zeta_{k_i} - \epsilon_{it})^2}{2(1-\alpha)}\right\}\pi(\tau).$$

Assume τ follows the truncated normal prior $\tau \sim N(0, 10)\mathbf{1}(\tau > 0)$. We then draw τ from the posterior distribution

$$N\left(\frac{\sum_{i=1}^{n}\sum_{t=0}^{T}\eta_{it}\zeta_{k_i}}{\frac{\sum_{i=1}^{n}\sum_{t=0}^{T}\zeta_{k_i}^2}{1-\alpha}+\frac{1}{10}}, \frac{1}{\frac{\sum_{i=1}^{n}\sum_{t=0}^{T}\zeta_{k_i}^2}{1-\alpha}+\frac{1}{10}}\right)\mathbf{1}(\tau > 0),$$

where $\eta_{it} = w_{it} - z_i\boldsymbol{\beta}_{1t} - \mathbf{x}_{1i}'\boldsymbol{\beta}_{2t} - \epsilon_{it}$.

Step (vi). The full conditional distribution of ζ_k is given by

$$\pi(\zeta_k|\mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\epsilon}, \alpha, \tau, \rho, D_{\text{obs}}) \propto \exp\left\{-\frac{\sum_{\{i|k_i=k\}}\sum_{t=0}^{T}(w_{it} - z_i\boldsymbol{\beta}_{1t} - \mathbf{x}_{1i}'\boldsymbol{\beta}_{2t} - \tau\zeta_{k_i} - \epsilon_{it})^2}{2(1-\alpha)}\right\}\exp\left(-\frac{\sum_{\{i|k_i=k\}}\sum_{t=0}^{T}\zeta_{k_i}^2}{2}\right).$$

We then draw ζ_k from a $N\left(\frac{\sum_{\{i|k_i=k\}}\sum_{t=0}^{T}\eta_{it}\frac{\tau}{1-\alpha}}{n_k(T+1)\frac{\tau^2}{1-\alpha}+n_k(T+1)}, \frac{1}{n_k(T+1)\frac{\tau^2}{1-\alpha}+n_k(T+1)}\right)$ distribution for $k = 1, \ldots, 16$, where n_k is the total number of patients in the k^{th} health district, i.e., $n_k = \sum_{\{i|k_i=k\}} 1$.

4.4 Bayesian Model Assessment

It is of great practical interest to try and assess whether the missingness is ignorable or nonignorable. In this section, several Bayesian model assessment criteria are considered, namely, the DIC relating to the missing data model ($\text{DIC}_{\mathbf{R}|\mathbf{y}}$)(Yao *et al.*, 2015; Mason *et al.*, 2012), and the LMPL relating to the missing data model ($\text{LPML}_{\mathbf{R}|\mathbf{y}}$) (Zhang *et al.*, 2014).

Since our focus is on the missing data mechanism, these criteria are applied only to the distribution of the missing data indicators. Both criteria are computationally attractive, and can be implemented with any types of priors, i.e., informative, noninformative, or even improper priors.

$\text{DIC}_{\mathbf{R}|\mathbf{y}}$. Let $\boldsymbol{\psi} = (\boldsymbol{\gamma}, \mathbf{y}_{\text{mis}})$ denote the vector of the missing data model parameters of interest, where we view $\mathbf{y}_{\text{mis}}$ as nuisance parameters. For the missing model in (3.8), $D(\boldsymbol{\psi}) = -2\sum_{i=0}^{n}\sum_{t=0}^{T}[r_{it}\eta_{it}^{r} - \log(1+\exp(\eta_{it}^{r}))]$. For computing $D(\overline{\boldsymbol{\psi}})$, we need to estimate several discrete parameters such as the binary response $\mathbf{y}_{\text{mis}}$. The posterior mean of $\mathbf{y}_{\text{mis}}$, which is no longer binary, may not be a desirable estimate to be applied in the $\text{DIC}_{\mathbf{R}|\mathbf{y}}$ formula. Instead, we may use the posterior mode, which maintains the binary nature of these parameters. Another possible choice given in Huang *et al.* (2005) is that we apply the linear predictor η_{it}^{r} directly to the $\text{DIC}_{\mathbf{R}|\mathbf{y}}$ formula. Therefore, we have $\text{DIC}_{\mathbf{R}|\mathbf{y}} = D(\overline{\boldsymbol{\eta}^{r}}) + 2p_D$, where $\overline{\eta_{it}^{r}} = E[z_i\gamma_{1t} + \mathbf{x}_{2i}'\boldsymbol{\gamma}_{2t} + g(\mathbf{R}_{it-1}, \boldsymbol{\gamma}_{3t}) + h(\mathbf{y}_{it}, \boldsymbol{\gamma}_{4t})|D_{\text{obs}}]$, $p_D = \overline{D(\boldsymbol{\psi})} - D(\overline{\boldsymbol{\psi}})$ is the effective number of parameters in the model, and $\overline{D(\boldsymbol{\psi})} = E[D(\boldsymbol{\psi})|D_{\text{obs}}]$. This modification is appropriate since the model for the missing data indicators depends on $\boldsymbol{\psi}$ only through the linear predictor $\boldsymbol{\eta}^{r}$. Moreover, with the introduction of $\boldsymbol{\eta}^{r}$ in the computation of $\text{DIC}_{\mathbf{R}|\mathbf{y}}$, we no longer need to worry about the discreteness of the parameters since $\boldsymbol{\eta}^{r}$ is always continuous. Similar to the traditional DIC, the model with the smallest $\text{DIC}_{\mathbf{R}|\mathbf{y}}$ value is the most optimal among all the models under consideration.

$\text{LPML}_{\mathbf{R}|\mathbf{y}}$. To assess the missing data mechanism, we adopt the conditional LPML (Hanson *et al.*, 2011), where the pseudomarginal probability, i.e., $\prod_{i=1}^{n} P(\mathbf{R}_{iT}|\mathbf{y}_i, z_i, \mathbf{x}_i, \boldsymbol{\gamma})$, is used to quantify the model's predictive ability. Let $D_{\text{obs}}^{(-i*)} = \{\mathbf{R}_{jT}, j = 1, \ldots, i-1, i+1, \ldots, n\} \cup \{(\mathbf{y}_{j,\text{obs}}, z_j, \mathbf{x}_j), j = 1, \ldots, n\}$ denote the observed data with $\mathbf{R}_{iT}$ deleted. Let $\boldsymbol{\psi}_1 = (\boldsymbol{\beta}, \tau, \boldsymbol{\zeta}, \alpha, \rho)$, and $\boldsymbol{\psi} = (\boldsymbol{\psi}_1, \boldsymbol{\gamma})$. Then we have

$$\pi(\boldsymbol{\psi}, \mathbf{y}_{\text{mis}}, \boldsymbol{\epsilon}|D_{\text{obs}}^{(-i*)}) \propto \left\{\prod_{j=1}^{n} f_y(\mathbf{y}_j|\boldsymbol{\psi}, z_j, \mathbf{x}_j, \boldsymbol{\epsilon}_j) f(\boldsymbol{\epsilon}_j|\alpha, \rho)\right\} \times \prod_{j\neq i} f_{\mathbf{R}|\mathbf{y}}(\mathbf{R}_{jT}|\boldsymbol{\gamma}, \mathbf{y}_j, z_j, \mathbf{x}_j)\pi(\boldsymbol{\psi}).$$

The simplified conditional predictive ordinate CPO_i (Chen *et al.*, 2000; Hanson *et al.*,

2011) can be written as

$$\begin{aligned}
\mathrm{CPO}_i &= \int \sum_{y_{i,\mathrm{mis}}} f_{\mathbf{R}|\mathbf{y}}(\mathbf{R}_{iT}|\boldsymbol{\gamma}, \mathbf{y}_i, z_i, \mathbf{x}_i)\pi(\boldsymbol{\psi}, \mathbf{y}_{\mathrm{mis}}, \boldsymbol{\epsilon}|D_{\mathrm{obs}}^{(-i*)})d\boldsymbol{\epsilon}d\boldsymbol{\psi} \\
&= \frac{1}{\int \sum_{y_{\mathrm{mis}}} \frac{1}{f_{\mathbf{R}|\mathbf{y}}(\mathbf{R}_{iT}|\boldsymbol{\gamma}, \mathbf{y}_i, z_i, \mathbf{x}_i)}\pi(\boldsymbol{\psi}, \mathbf{y}_{\mathrm{mis}}, \boldsymbol{\epsilon}|D_{\mathrm{obs}})d\boldsymbol{\epsilon}d\boldsymbol{\psi}},
\end{aligned}$$

and the logarithm of the pseudomarginal likelihood is given by

$$\mathrm{LPML}_{\mathbf{R}|\mathbf{y}} = \sum_{i=1}^{n} \log(\mathrm{CPO}_i).$$

Let $\{(\boldsymbol{\psi}_b, \mathbf{y}_{\mathrm{mis},b}, \boldsymbol{\epsilon}_b), b = 1, \ldots, B\}$ denote a Gibbs sample of $(\boldsymbol{\psi}, \mathbf{y}_{\mathrm{mis}}, \boldsymbol{\epsilon})$ from (4.6) and let b represent the b^{th} iteration. A Monte Carlo estimate of CPO_i is given by

$$\mathrm{CPO}_i = \left(\frac{1}{B}\sum_{b=1}^{B} \frac{1}{f_{\mathbf{R}|\mathbf{y}}(\mathbf{R}_{iT}|\mathbf{y}_{i,\mathrm{obs}}, z_i, \mathbf{x}_i, \boldsymbol{\psi}_b, \mathbf{y}_{i,\mathrm{mis},b}, \boldsymbol{\epsilon}_{i,b})}\right)^{-1}.$$

Similar to the conventional LPML, a large value of $\mathrm{LPML}_{\mathbf{R}|\mathbf{y}}$ indicates a more favorable model.

5 A Simulation Study

In this section, we conduct a simulation study to investigate the empirical performance of the proposed method. In the data generation, we first generated $n = 2000$ baseline covariates as follows: $x_{1i} \sim N(0, 1)$, $x_{2i}|x_{1i} \sim \mathrm{Bernoulli}(1/(1 + \exp(-0.2 - 0.2x_{1i})))$, and the intervention indicator $z_i \sim \mathrm{Bernoulli}(0.5)$. Similar to the HIV prevention behavioral data, we set the total number of visits equal 4. Let $\boldsymbol{\epsilon}^*$ in (3.1) follow a $N(\mathbf{0}, \sigma^2\Sigma)$ distribution, where $\sigma^2 = 2$ ($\alpha \approx 0.667$) and Σ is a 4×4 AR(1) correlation matrix with $\rho = 0.8$. The longitudinal binary response variable y_{it} was generated from a Bernoulli distribution with the probability of $y_{it} = 1$ given by

$$P(y_{it} = 1|z_i, x_{1i}, x_{2i}, \boldsymbol{\beta}_t^*, \epsilon_{it}^*) = \Phi(\beta_{0t}^* + x_{1i}\beta_{1t}^* + x_{2i}\beta_{2t}^* + z_i\beta_{3t}^* + \epsilon_{it}^*),$$

where $\boldsymbol{\beta}_t^* = (\beta_{0t}^*, \beta_{1t}^*, \beta_{2t}^*, \beta_{3t}^*)'$ for $t = 0, 1, 2, 3$. To reproduce the longitudinal binary response data pattern of the HIV prevention behavioral data, we set

$$\begin{pmatrix} \boldsymbol{\beta}_0^{*\prime} \\ \boldsymbol{\beta}_1^{*\prime} \\ \boldsymbol{\beta}_2^{*\prime} \\ \boldsymbol{\beta}_3^{*\prime} \end{pmatrix} = \begin{pmatrix} -1.0 & 0.5 & 1.0 & 0.4 \\ -1.0 & 0.5 & 1.0 & -0.2 \\ -1.0 & 0.5 & 1.0 & -0.4 \\ -1.0 & 0.5 & 1.0 & -0.6 \end{pmatrix}. \tag{5.1}$$

We then generated the missing data indicator $R_{it} \sim \text{Bernoulli}(P_{it})$, where P_{it} is given by

$$\text{logit}(P_{it}) = \gamma_{0t} + x_{1i}\gamma_{1t} + x_{2i}\gamma_{2t} + z_i\gamma_{3t} + \sum_{j=0}^{t-1} R_{ij}\gamma_{4t} + y_{it-1}\gamma_{5t} + y_{it}\gamma_{6t}. \tag{5.2}$$

The missing data mechanism is, therefore, nonignorably missing since P_{it} in (5.2) depends on the unobserved data y_{it-1} and y_{it} when $R_{i,t-1} = R_{it} = 1$. Let $\boldsymbol{\gamma}_t = (\gamma_{0t}, \gamma_{1t}, \gamma_{2t}, \gamma_{3t}, \gamma_{4t}, \gamma_{5t}, \gamma_{6t})'$ for $t = 0, 1, 2, 3$. We set

$$\begin{pmatrix} \boldsymbol{\gamma}_0{}' \\ \boldsymbol{\gamma}_1{}' \\ \boldsymbol{\gamma}_2{}' \\ \boldsymbol{\gamma}_3{}' \end{pmatrix} = \begin{pmatrix} -2.50 & 0.50 & -0.50 & -0.50 & 0.00 & 0.00 & 0.00 \\ -2.00 & 0.50 & -0.50 & -0.25 & -0.25 & 0.50 & 0.40 \\ -2.80 & 0.50 & -0.50 & 0.25 & -0.60 & 1.30 & 1.70 \\ -2.80 & 0.50 & -0.50 & 0.50 & 0.60 & -0.50 & 1.70 \end{pmatrix}. \tag{5.3}$$

Under this setting, the average missingness percentages across the 250 simulated data sets were 5.37%, 10.52%, 11.94%, and 14.18% at $t = 0, 1, 2, 3$, respectively.

To further examine the performance of the proposed method, we also considered another scenario, in which the missingness percentage of the last visit ($t = 3$) was set to 47.14% and the missingness percentages at the other time points remained the same. This was achieved by setting γ_{03} in (5.3) equal -0.50. In the simulation, we assigned the true values to the initial values for each parameter. After discarding the first 500 iterations of the sampler, we used the subsequent 5,000 iterations for computing the posterior summaries.

We fit both the ignorable and nonignorable models to the simulated data generated from the nonignorable model. For the ignorable model, we set γ_{5t} and γ_{6t} in (5.2) equal 0 so that P_{it} depends only on the intervention indicator, the covariates $\mathbf{x}_2$, as well as the cumulative number of missing visits, which all were observed. For the nonignorable model, we considered Jeffreys prior for $\boldsymbol{\gamma}_t$ in (4.5), as well as a $N(0, \sigma^2_{prior})$ prior for γ_{6t}, where $\sigma^2_{prior} = 1, 2, \ldots, 10$.

When the missingness percentage was low (similar to the real data), the median (IQR) of $\text{DIC}_{\mathbf{R}|\mathbf{y}}$ under the ignorable model was 4562.49 (4490.64, 4641.60). The nonignorable model with a N(0, 10) prior had the smallest median value of $\text{DIC}_{\mathbf{R}|\mathbf{y}}$ (4473.76 (4381.28, 4465.02)). The median (IQR) of $\text{LPML}_{\mathbf{R}|\mathbf{y}}$ under the ignorable model was -2281.40 (-2320.90, -2245.39). Among all the normal priors, the nonignorable model with a N(0, 6) prior had the largest median value of $\text{LPML}_{\mathbf{R}|\mathbf{y}}$ (-2273.04 (-2313.26, -2234.85)), and the nonignorable model with the Jeffreys prior had the largest value (-2272.85 (-2311.38, -2235.87)) of $\text{LPML}_{\mathbf{R}|\mathbf{y}}$ among all the models under consideration.

For the high missingness percentage scenario (47.14% missing at the last visit), the median (IQR) of $\text{DIC}_{\mathbf{R}|\mathbf{y}}$ under the ignorable model was 5673.07 (5605.66, 5741.60). The nonignorable model with a N(0, 10) prior still had the smallest median value of $\text{DIC}_{\mathbf{R}|\mathbf{y}}$ (5559.20 (5471.43, 5644.64)). The median (IQR) of $\text{LPML}_{\mathbf{R}|\mathbf{y}}$ under the ignorable model was -2836.63 (-2870.99, -2802.92). Among all the normal priors, the nonignorable model with a N(0, 8) prior had the largest median value of $\text{LPML}_{\mathbf{R}|\mathbf{y}}$ (-2816.79 (-2858.90, -2781.31)), and the nonignorable model with the Jeffreys prior had the largest value (-2815.01 (-2849.76, -2780.99)) among all the models under consideration.

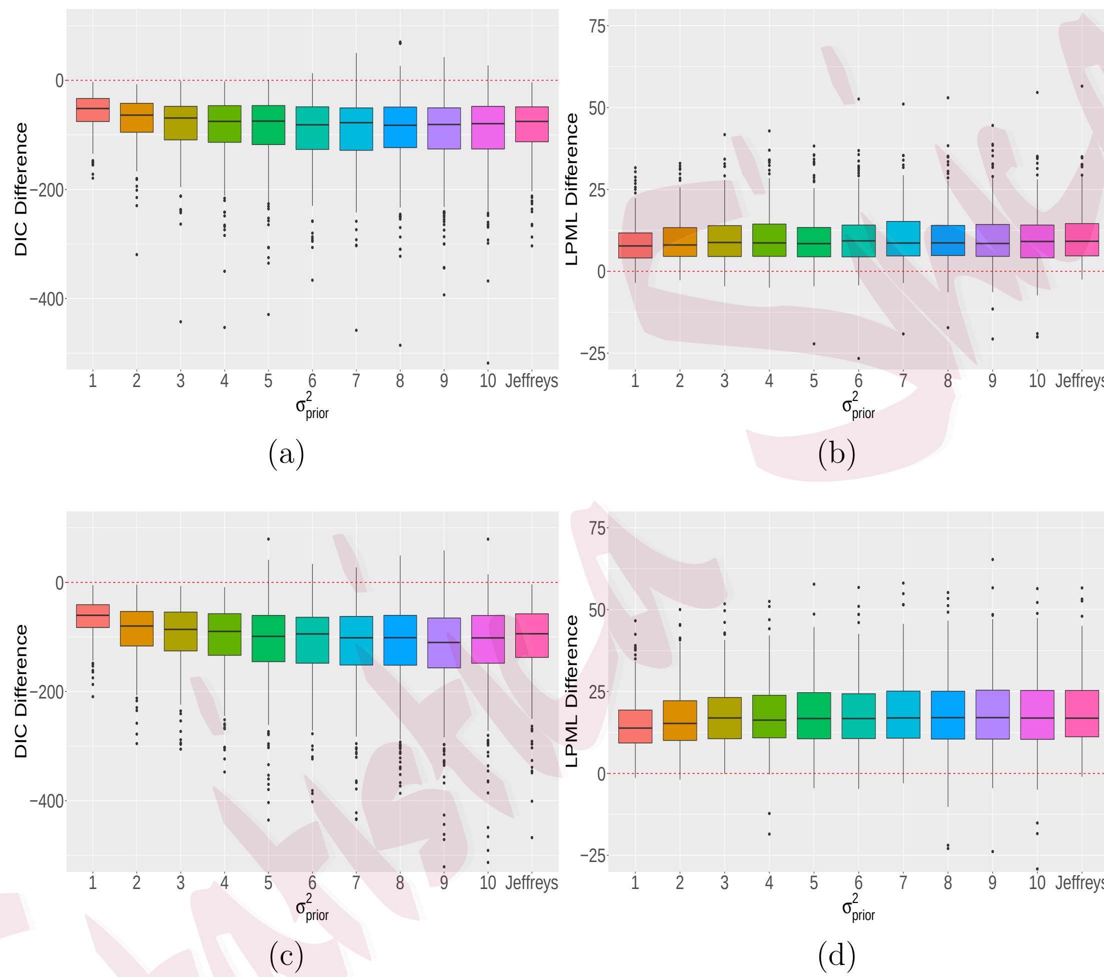

Figure 2: Plots of the DIC differences (a) and the LPML differences (b) when the missingness percentages were 5.37%, 10.52%, 11.94%, and 14.18%; and plots of the DIC differences (c) and the LPML differences (d) when the missingness percentages were 5.37%, 10.52%, 11.94%, and 47.14%.

Let the "DIC Difference" be the $\text{DIC}_{\mathbf{R}|\mathbf{y}}$ under the nonignorable model minus the $\text{DIC}_{\mathbf{R}|\mathbf{y}}$ under the ignorable model. Similarly, let the "LPML Difference" be the $\text{LPML}_{\mathbf{R}|\mathbf{y}}$ under the nonignorable model minus the $\text{LPML}_{\mathbf{R}|\mathbf{y}}$ under the ignorable model. Figure 2 shows the plots of the DIC differences and the LPML differences versus different priors ($N(0, \sigma^2_{prior})$'s or Jeffreys) specified under the nonignorable model under the two scenarios with different missingness percentages. From Figure 2, we see that (i) the DIC differences first decrease and then slightly increase as σ^2_{prior} increases (Figure 2(a) and Figure 2(c)); and (ii) the LPML differences first increase and then slightly decrease as σ^2_{prior} increases (Figure 2(b) and Figure 2(d)) under both scenarios. Based on Figure 2(a) and Figure 2(b), when the missingness percentage is low, the nonignorable model with $N(0, 6)$ seemed to have the best relative performance. For the high missingness percentage case (Figure 2(c) and Figure 2(d)), the nonignorable model with $N(0, 9)$ tended to perform comparatively better. Moreover, all of the boxes for the "DIC Difference" were below 0, and all of the boxes for the "LPML Difference" were above 0, indicating that both $\text{DIC}_{\mathbf{R}|\mathbf{y}}$ and $\text{LPML}_{\mathbf{R}|\mathbf{y}}$ were in favor of the nonignorable model over the ignorable model. Also, as the missingness percentage increases, the boxes for both "DIC Difference" and "LPML Difference" became further away from the horizontal line ($y = 0$), implying that the power of the two criteria increased as the missingness percentage increased.

Tables 2 and 3 show the true value of the parameter (True), the posterior mean (Est), the standard deviation of the estimate (SD), the average of the posterior standard deviations (SE), the root of the mean squared error of the posterior mean (RMSE), and the coverage probability (CP) of the 95% highest posterior density (HPD) interval for each parameter across 250 simulations under the nonignorable models with the $N(0, 6)$ prior and Jeffreys prior for the low missingness percentage case and the nonignorable models with the $N(0, 8)$ prior and Jeffreys prior for the high missingness percentage case. We see from these tables that (i) all of the posterior estimates were close to the true values; (ii) SDs, SEs, and RMSEs were close to each other; and (iii) CPs for most of the parameters were approximately 95%, except for some of the γ_{5t} and γ_{6t}. The posterior estimates under the other priors are given in Tables S1 and S2 in the Supplemental Materials. From these tables, we see that the posterior estimates were quite robust to the specification of the $N(0, \sigma^2_{prior})$ prior under the nonignorable model.

6 Analysis of the HIV Prevention Behavioral Data

In this section, we carry out a detailed analysis of the HIV prevention behavioral data discussed in Section 2. The baseline covariates in the response model and missing data mechanism include Gender (1=female), City (1=Lives in city or township), Cohabit (1=Cohabitates with sex partner), Counselor (1=Meets with a counselor at least every 3 months), Drink (1=Reported drinking alcohol weekly or more frequently), and Age. Except for Age, which is continuous, all other covariates are binary. Due to the rare events of Drink in the "missing" group of patients, the Drink covariate is not identifiable, and is therefore excluded in the missing data mechanism. For the missing data mechanism, we also

Table 2: Posterior Summaries under the Nonignorable Model with a N(0, 6) Prior and Jeffreys Prior When the Missingness Percentages Were 5.37%, 10.52%, 11.94%, and 14.18%

		N(0, 6) Prior					Jeffreys Prior				
	TRUE	EST	SD	SE	RMSE	CP	EST	SD	SE	RMSE	CP
t=0											
β^*_{00}	-1.000	-1.008	0.134	0.125	0.125	0.976	-1.011	0.135	0.125	0.125	0.972
β^*_{10}	0.500	0.505	0.068	0.068	0.068	0.960	0.506	0.069	0.069	0.070	0.960
β^*_{20}	1.000	1.002	0.132	0.129	0.129	0.952	1.006	0.133	0.129	0.129	0.952
β^*_{30}	0.400	0.402	0.110	0.098	0.098	0.976	0.403	0.110	0.099	0.098	0.980
γ_{00}	-2.500	-2.669	0.355	0.372	0.408	0.960	-2.666	0.354	0.495	0.521	0.960
γ_{10}	0.500	0.502	0.125	0.120	0.120	0.960	0.499	0.125	0.120	0.119	0.964
γ_{20}	-0.500	-0.485	0.250	0.245	0.245	0.960	-0.480	0.248	0.242	0.242	0.956
γ_{30}	-0.500	-0.499	0.217	0.204	0.203	0.968	-0.493	0.215	0.200	0.200	0.968
γ_{60}	0.000	-0.011	0.845	0.804	0.803	0.972	-0.004	0.878	0.921	0.919	0.960
t=1											
β^*_{01}	-1.000	-0.994	0.165	0.179	0.179	0.924	-1.002	0.163	0.169	0.169	0.940
β^*_{11}	0.500	0.499	0.073	0.068	0.068	0.980	0.500	0.073	0.069	0.069	0.960
β^*_{21}	1.000	0.982	0.143	0.145	0.146	0.940	0.988	0.143	0.140	0.140	0.932
β^*_{31}	-0.200	-0.195	0.110	0.104	0.104	0.944	-0.196	0.110	0.105	0.105	0.940
γ_{01}	-2.000	-2.173	0.340	0.358	0.397	0.956	-2.130	0.306	0.359	0.381	0.960
γ_{11}	0.500	0.505	0.094	0.096	0.096	0.924	0.504	0.092	0.097	0.097	0.920
γ_{21}	-0.500	-0.513	0.191	0.201	0.201	0.932	-0.508	0.188	0.193	0.192	0.940
γ_{31}	-0.250	-0.262	0.163	0.157	0.157	0.964	-0.262	0.162	0.153	0.153	0.968
γ_{41}	0.400	0.390	0.295	0.301	0.300	0.944	0.375	0.292	0.300	0.301	0.944
γ_{51}	-0.250	-0.257	0.297	0.297	0.297	0.924	-0.246	0.290	0.288	0.287	0.940
γ_{61}	0.500	0.550	0.874	0.918	0.917	0.932	0.495	0.848	0.937	0.935	0.956
t=2											
β^*_{02}	-1.000	-1.014	0.152	0.162	0.162	0.952	-1.024	0.152	0.156	0.158	0.956
β^*_{12}	0.500	0.497	0.071	0.067	0.067	0.964	0.498	0.072	0.068	0.068	0.960
β^*_{22}	1.000	1.004	0.145	0.141	0.141	0.956	1.012	0.145	0.138	0.138	0.960
β^*_{32}	-0.400	-0.395	0.114	0.110	0.110	0.944	-0.398	0.115	0.110	0.110	0.944
γ_{02}	-2.800	-2.952	0.323	0.382	0.411	0.932	-2.899	0.301	0.348	0.361	0.920
γ_{12}	0.500	0.502	0.090	0.097	0.097	0.956	0.499	0.089	0.096	0.096	0.944
γ_{22}	-0.500	-0.523	0.188	0.181	0.182	0.968	-0.515	0.186	0.177	0.177	0.960
γ_{32}	0.250	0.268	0.165	0.179	0.179	0.932	0.262	0.163	0.175	0.175	0.932
γ_{42}	1.700	1.761	0.180	0.195	0.204	0.936	1.738	0.176	0.188	0.191	0.944
γ_{52}	-0.600	-0.616	0.270	0.316	0.316	0.916	-0.602	0.267	0.303	0.303	0.904
γ_{62}	1.300	1.383	0.617	0.722	0.725	0.920	1.335	0.585	0.662	0.661	0.940
t=3											
β^*_{03}	-1.000	-1.004	0.142	0.142	0.141	0.948	-1.007	0.143	0.142	0.141	0.952
β^*_{13}	0.500	0.502	0.076	0.080	0.080	0.936	0.504	0.077	0.081	0.081	0.936
β^*_{23}	1.000	1.006	0.141	0.131	0.131	0.956	1.010	0.142	0.132	0.132	0.956
β^*_{33}	-0.600	-0.604	0.122	0.121	0.121	0.948	-0.606	0.123	0.121	0.121	0.948
γ_{03}	-2.800	-2.892	0.189	0.202	0.221	0.932	-2.865	0.186	0.197	0.207	0.940
γ_{13}	0.500	0.500	0.092	0.098	0.098	0.940	0.496	0.091	0.096	0.096	0.936
γ_{23}	-0.500	-0.499	0.174	0.171	0.171	0.956	-0.496	0.173	0.170	0.170	0.952
γ_{33}	0.500	0.518	0.165	0.173	0.174	0.936	0.512	0.164	0.171	0.171	0.940
γ_{43}	1.700	1.748	0.119	0.122	0.131	0.944	1.736	0.117	0.121	0.126	0.968
γ_{53}	0.600	0.580	0.261	0.255	0.255	0.948	0.575	0.258	0.250	0.250	0.952
γ_{63}	-0.500	-0.495	0.562	0.595	0.594	0.940	-0.485	0.548	0.581	0.580	0.916
ρ	0.800	0.795	0.038	0.036	0.037	0.948	0.794	0.038	0.036	0.036	0.948
α	0.667	0.662	0.046	0.044	0.044	0.956	0.663	0.046	0.044	0.044	0.956

Table 3: Posterior Summaries under the Nonignorable Model with a N(0, 8) Prior and Jeffreys Prior When the Missingness Percentages Were 5.37%, 10.52%, 11.94%, and 47.14%

		N(0, 8) Prior					Jeffreys Prior				
	TRUE	EST	SD	SE	RMSE	CP	EST	SD	SE	RMSE	CP
t=0											
β^*_{00}	-1.000	-1.004	0.148	0.131	0.131	0.972	-1.012	0.146	0.134	0.134	0.972
β^*_{10}	0.500	0.504	0.073	0.071	0.071	0.960	0.506	0.074	0.073	0.073	0.968
β^*_{20}	1.000	1.000	0.143	0.135	0.135	0.952	1.006	0.143	0.137	0.137	0.968
β^*_{30}	0.400	0.400	0.113	0.101	0.100	0.976	0.403	0.113	0.101	0.101	0.980
γ_{00}	-2.500	-2.715	0.442	0.417	0.468	0.960	-2.648	0.348	0.411	0.436	0.960
γ_{10}	0.500	0.499	0.128	0.118	0.118	0.972	0.501	0.125	0.118	0.118	0.972
γ_{20}	-0.500	-0.490	0.255	0.247	0.246	0.952	-0.476	0.248	0.239	0.240	0.968
γ_{30}	-0.500	-0.502	0.218	0.204	0.203	0.972	-0.492	0.215	0.202	0.202	0.972
γ_{60}	0.000	0.041	0.960	0.835	0.834	0.964	-0.047	0.892	0.877	0.877	0.972
t=1											
β^*_{01}	-1.000	-0.982	0.182	0.192	0.193	0.924	-0.997	0.178	0.190	0.189	0.920
β^*_{11}	0.500	0.499	0.078	0.074	0.074	0.972	0.500	0.078	0.076	0.076	0.956
β^*_{21}	1.000	0.974	0.155	0.152	0.154	0.932	0.984	0.155	0.153	0.154	0.940
β^*_{31}	-0.200	-0.197	0.111	0.105	0.105	0.944	-0.196	0.112	0.104	0.104	0.952
γ_{01}	-2.000	-2.258	0.429	0.485	0.549	0.952	-2.173	0.346	0.395	0.430	0.952
γ_{11}	0.500	0.501	0.096	0.100	0.100	0.912	0.503	0.094	0.100	0.100	0.916
γ_{21}	-0.500	-0.525	0.196	0.208	0.209	0.936	-0.512	0.192	0.197	0.197	0.952
γ_{31}	-0.250	-0.257	0.165	0.158	0.158	0.964	-0.260	0.163	0.155	0.155	0.968
γ_{41}	0.400	0.396	0.300	0.305	0.304	0.948	0.377	0.295	0.302	0.302	0.944
γ_{51}	-0.250	-0.278	0.310	0.324	0.324	0.924	-0.254	0.299	0.317	0.316	0.936
γ_{61}	0.500	0.644	1.019	1.127	1.134	0.928	0.507	0.961	1.124	1.122	0.908
t=2											
β^*_{02}	-1.000	-1.010	0.169	0.167	0.167	0.948	-1.025	0.167	0.165	0.167	0.936
β^*_{12}	0.500	0.496	0.077	0.071	0.071	0.960	0.496	0.078	0.075	0.075	0.956
β^*_{22}	1.000	0.999	0.156	0.149	0.149	0.968	1.010	0.157	0.150	0.150	0.948
β^*_{32}	-0.400	-0.395	0.117	0.112	0.112	0.948	-0.397	0.118	0.113	0.113	0.952
γ_{02}	-2.800	-2.987	0.361	0.437	0.475	0.924	-2.920	0.331	0.402	0.418	0.924
γ_{12}	0.500	0.501	0.092	0.101	0.101	0.932	0.500	0.090	0.098	0.098	0.940
γ_{22}	-0.500	-0.527	0.195	0.186	0.187	0.964	-0.513	0.191	0.182	0.182	0.960
γ_{32}	0.250	0.268	0.168	0.181	0.181	0.928	0.260	0.165	0.178	0.178	0.928
γ_{42}	1.700	1.772	0.185	0.199	0.211	0.948	1.746	0.179	0.188	0.193	0.944
γ_{52}	-0.600	-0.614	0.287	0.326	0.326	0.916	-0.589	0.282	0.324	0.323	0.912
γ_{62}	1.300	1.404	0.710	0.829	0.833	0.940	1.321	0.668	0.781	0.780	0.916
t=3											
β^*_{03}	-1.000	-0.970	0.219	0.242	0.243	0.904	-0.973	0.219	0.234	0.236	0.908
β^*_{13}	0.500	0.508	0.103	0.102	0.102	0.944	0.511	0.104	0.103	0.103	0.944
β^*_{23}	1.000	0.988	0.174	0.165	0.165	0.952	0.994	0.177	0.167	0.167	0.956
β^*_{33}	-0.600	-0.598	0.152	0.156	0.156	0.952	-0.599	0.153	0.157	0.157	0.948
γ_{03}	-0.500	-0.547	0.133	0.147	0.155	0.912	-0.545	0.132	0.139	0.146	0.936
γ_{13}	0.500	0.503	0.064	0.065	0.065	0.960	0.500	0.063	0.064	0.064	0.968
γ_{23}	-0.500	-0.504	0.118	0.127	0.127	0.924	-0.504	0.118	0.124	0.124	0.940
γ_{33}	0.500	0.511	0.109	0.115	0.115	0.936	0.509	0.109	0.113	0.113	0.948
γ_{43}	1.700	1.733	0.137	0.142	0.146	0.952	1.727	0.137	0.141	0.143	0.956
γ_{53}	0.600	0.578	0.188	0.203	0.204	0.952	0.573	0.187	0.199	0.200	0.940
γ_{63}	-0.500	-0.466	0.443	0.511	0.511	0.888	-0.452	0.438	0.480	0.482	0.916
ρ	0.800	0.796	0.044	0.041	0.041	0.948	0.796	0.044	0.041	0.041	0.952
α	0.667	0.658	0.052	0.048	0.049	0.964	0.660	0.052	0.049	0.049	0.968

consider covariates $\mathbf{y}_t$, and $\sum_{j=0}^{t-1} R_j$ at the t^{th} visit. For the HIV prevention behavioral data, we have $K = 16$ health districts and $T = 3$, where $t = 0$ denotes "baseline", and visits $t = 1$ to $t = 3$ correspond to the three follow-up visits at 6, 12, and 18 months. The continuous covariate Age was standardized for numerical stability in the posterior computations.

In all the Bayesian computations, we used 20,000 MCMC samples, which were taken from every fifth iteration, after a burn-in of 10,000 iterations for each model to compute all posterior summaries, including posterior means (ESTs), posterior standard deviations (SDs), 95% HPD intervals, DIC, and LPML. The code was written in FORTRAN 95 using IMSL subroutines with double-precision accuracy. The convergence of the Gibbs sampler was checked by the R package "mcmcplots" using R version 3.3.0. Approximate convergence was reached after 10,000 iterations.

Table 4: Values of $\text{DIC}_{\mathbf{R}|\mathbf{y}}$ (p_D) and $\text{LPML}_{\mathbf{R}|\mathbf{y}}$ under Ignorable Missingness and Nonignorable Missingness with Various Priors for the HIV Prevention Behavioral Data

Fitted Model	p_D	$\text{DIC}_{\mathbf{R}\|\mathbf{y}}$	$\text{LPML}_{\mathbf{R}\|\mathbf{y}}$
Ignorable	30.85	4793.16	-2397.24
Nonignorable N(0, 1)	89.82	4769.73	-2398.26
Nonignorable N(0, 2)	107.06	4755.71	-2397.44
Nonignorable N(0, 3)	114.95	4757.82	-2397.86
Nonignorable N(0, 4)	112.99	4751.86	-2397.70
Nonignorable N(0, 5)	126.66	4748.78	-2397.28
Nonignorable N(0, 6)	132.95	4746.74	-2397.23
Nonignorable N(0, 7)	132.67	4747.22	-2397.23
Nonignorable N(0, 8)	132.94	**4737.61**	**-2396.32**
Nonignorable N(0, 9)	133.47	4745.62	-2397.29
Nonignorable N(0, 10)	140.61	4749.97	-2398.21
Nonignorable Jeffreys Prior	120.18	4750.08	**-2396.64**

We fit the ignorable and nonignorable models to the HIV prevention behavioral data. For the ignorable model, we simply set $h(\mathbf{y}_t, \boldsymbol{\gamma}_{4t}) = 0$ in (3.8). For the nonignorable model, we assumed that $h(\mathbf{y}_t, \boldsymbol{\gamma}_{4t}) = \gamma_{4t1} y_{t-1} + \gamma_{4t2} y_t$ in (3.8) and considered a $N(0, \sigma^2_{prior})$ prior for γ_{4t2} as well as Jeffreys prior for $\boldsymbol{\gamma}_t$ in (4.5). We specified uniform priors for all other parameters. We then computed DIC and LPML under the ignorable model, the nonignorable model using a $N(0, \sigma^2_{prior})$ prior, and the nonignorable model using Jeffreys prior. The values of DIC and LPML are shown in Table 4. As exhibited in Table 4, the effective number of parameters under the ignorable model ($p_D = 30.85$) was the smallest among all the models we considered, and approximately equal to the number of parameters. Under the nonignorable model with a $N(0, \sigma^2_{prior})$ prior, the effective number of parameters increased with σ^2_{prior}. Moreover, p_D under Jeffreys prior was midway between p_D under the $N(0,4)$ and $N(0,5)$ priors. We also see from Table 4 that (i) the DIC value was 4793.16 under the ignorable model; (ii) under the nonignorable model with a $N(0,\sigma^2_{prior})$ prior, the value of DIC first tended to decrease and then increase as σ^2_{prior} increased; (iii) the DIC attained the local minimum with DIC=4737.61 at $\sigma^2_{prior}=8$ among all the models under consideration (10 values of σ^2_{prior} and Jeffreys Prior). The results indicated by LPML were consistent with the results by the DIC criterion. The nonignorable model with a $N(0,8)$ prior had the largest value of LPML (LPML=-2396.32) among all the models under consideration. The nonignorable model with Jeffreys prior had the second largest value of LPML (LPML=-2396.64). These results indicate that for the HIV prevention behavioral data, the missing longitudinal binary responses were potentially nonignorably missing.

Tables 5-7 show the ESTs, SDs, and 95% HPD intervals under the ignorable model, the nonignorable model with the $N(0,8)$ prior, and the nonignorable model with Jeffreys prior. We define a posterior estimate to be "statistically significant at a significance level of 0.05" if the corresponding 95% HPD interval does not contain 0. Under the ignorable model, based on the posterior estimates of the intervention effect (z) in Table 5, the counseling intervention significantly reduced HIV risk behavior after 6-Month. The covariate Cohabit was always significant (at each visit), indicating that people who cohabitated with their primary sex partner were more likely to experience unprotected sex acts. Gender (at Baseline and 12-Month), Cohabit (at each visit), Counselor (at baseline, 6-Month, and 18-Month), and Drink (at 6-Month) all had significant positive posterior estimates, which means females, people visiting counselors more frequently, and people who drank more often tended to have more HIV behavior risks. Age (at each visit) had a strong negative effect on the HIV behavior risk, indicating that older people may have better knowledge of safe sexual behavior. For the missing data mechanism, the posterior estimates of Condition varied from negative to positive values as time progressed, indicating that people in the intervention arm tended to participate in the study at the very beginning and then became more likely to leave the study later. This behavior could possibly be explained by the conjecture that people who have already accumulated enough behavioral knowledge may consider it unnecessary to continue the risk prevention study. Females (at 6-Month, 12-Month and 18-Month) and older people (at 12-Month) were less likely to miss their visits, while people who lived in a city or town (18-Month) were likely to drop out at the last visit. Moreover, people who frequently skipped the previous visits had higher odds of missingness in the future, as indicated by the cumulative number of missing data indicators ($\sum_{j=0}^{t} R_j$).

The posterior estimates in Table 6 were similar to those given in Table 5. However, Gender (at 12-Month), which is a covariate in the response model, was significant with 95% HPD interval=(0.051, 0.636) under the ignorable model but not significant with 95% HPD interval=(-0.069, 0.525) under the nonignorable model with a $N(0,8)$ prior. Similarly, Age (at 12-Month), which is a covariate in the missing data mechanism, was significant with 95% HPD interval=(-0.309, -0.019) in the ignorable case but not significant with 95% HPD interval=(-0.272, 0.072) in Table 6. However, the covariates in the missing data mechanism, y_1 (95% HPD interval=(-1.239, -0.015)) and y_2 (95% HPD interval=(0.035, 2.822)) at 12-Month, and y_2 at 18-Month (95% HPD interval=(0.043, 1.169)) were all significant, indicating that missingness of the binary responses may be nonignorable. This result was consistent with the DIC and LPML.

In addition, the posterior standard deviations in Table 6 were similar to those given in Table 5 in the binary model. For the covariates in the missing data mechanism shared in both the ignorable and nonignorable models, the posterior standard deviations in Table 6 in the missing data mechanism, were generally larger than those given in Table 5. The standard deviation of γ_{4t2} corresponding to the missing response covariate y_t increased as σ^2_{prior} increased, implying that γ_{4t2} could not be estimated under an improper uniform prior. It is apparent that the posterior estimates under the nonignorable model were different than those under the ignorable model. The posterior estimates under the nonignorable model with Jeffreys prior (in Table 7) were similar to those under the nonignorable model with a $N(0, 8)$ prior (in Table 6) for both the binary response model and missing data mechanism, except that the standard deviations for the missing data mechanism in Table 7 were slightly smaller. The posterior estimates of ρ, α and τ were similar under the three models.

7 Discussion

In this paper, we developed Bayesian methods for resolving the challenges in estimation and Bayesian computation of the longitudinal binary probit model with nonignorably missing response data. An alternative longitudinal binary probit model is given by Chib and Greenberg (1998), in which identifiability of the variance of random effects in (3.3) is avoided by setting σ^2 equal to 1. However, this approach requires integrating out the high-dimensional truncated multivariate normal latent variables $\mathbf{w}$ when sampling the missing responses. For the missing data mechanism in (3.8), one may modify the model by relaxing the linear assumptions on g and h. Even in the same formulation, the model can be extended by including interaction terms between treatment and other covariates. If the missing data mechanism has too many covariates, however, it may lead to the problem of overfitting and may require a larger dataset to be identifiable. Thus, it is more desirable to develop a simple and identifiable model that leads to a good fit.

In this paper, we construct Jeffreys prior in (4.5) using a subset of the data, which are completely observed. Based on our simulation study in Section 5, the Jeffreys prior in (4.5) does yield quite good frequentist properties of the posterior estimates. As empirically

Table 5: Posterior Summaries under the Ignorable Model for the HIV Prevention Behavioral Data

	Binary Response Model				Missing Data Mechanism		
	EST	SD	95% HPD Interval		EST	SD	95% HPD Interval
Baseline				**Baseline**			
Intercept	-0.694	0.196	(-1.063, -0.291)	Intercept	-3.490	0.411	(-4.296, -2.689)
Gender	0.379	0.132	(0.114, 0.634)	Gender	0.115	0.237	(-0.336, 0.591)
City	0.123	0.157	(-0.186, 0.432)	City	-0.334	0.328	(-0.986, 0.290)
Cohabit	0.720	0.140	(0.455, 1.002)	Cohabit	0.229	0.227	(-0.242, 0.654)
Counselor	0.433	0.158	(0.127, 0.749)	Counselor	0.664	0.367	(-0.057, 1.380)
Drink	0.435	0.350	(-0.243, 1.129)	Age	0.083	0.111	(-0.129, 0.305)
Age	-0.372	0.073	(-0.516, -0.234)	—	—	—	—
6-Month				**6-Month**			
Intercept	-1.756	0.268	(-2.274, -1.246)	Intercept	-2.101	0.227	(-2.537, -1.651)
Gender	0.151	0.137	(-0.124, 0.415)	Gender	-0.397	0.149	(-0.690, -0.107)
City	0.112	0.167	(-0.211, 0.445)	City	0.030	0.183	(-0.314, 0.395)
Cohabit	0.638	0.145	(0.354, 0.923)	Cohabit	0.220	0.144	(-0.065, 0.500)
Counselor	0.574	0.179	(0.227, 0.917)	Counselor	0.274	0.196	(-0.080, 0.691)
Drink	0.987	0.372	(0.273, 1.726)	Age	-0.101	0.075	(-0.252, 0.042)
Age	-0.463	0.083	(-0.630, -0.310)	R_0	0.364	0.302	(-0.234, 0.949)
12-Month				**12-Month**			
Intercept	-1.811	0.281	(-2.371, -1.289)	Intercept	-1.953	0.211	(-2.351, -1.522)
Gender	0.331	0.150	(0.051, 0.636)	Gender	-0.482	0.144	(-0.760, -0.199)
City	-0.005	0.173	(-0.337, 0.345)	City	-0.117	0.183	(-0.465, 0.249)
Cohabit	0.638	0.151	(0.344, 0.935)	Cohabit	-0.107	0.141	(-0.385, 0.167)
Counselor	0.275	0.182	(-0.078, 0.627)	Counselor	-0.249	0.175	(-0.591, 0.094)
Drink	0.594	0.366	(-0.131, 1.293)	Age	-0.160	0.074	(-0.309, -0.019)
Age	-0.488	0.088	(-0.662, -0.323)	$\sum_{j=0}^{1} R_j$	1.644	0.140	(1.369, 1.918)
18-Month				**18-Month**			
Intercept	-1.750	0.275	(-2.273, -1.219)	Intercept	-2.641	0.238	(-3.111, -2.187)
Gender	0.241	0.148	(-0.046, 0.534)	Gender	-0.381	0.153	(-0.676, -0.079)
City	-0.143	0.182	(-0.510, 0.201)	City	0.403	0.181	(0.051, 0.763)
Cohabit	0.493	0.146	(0.209, 0.786)	Cohabit	0.081	0.149	(-0.212, 0.370)
Counselor	0.408	0.185	(0.047, 0.771)	Counselor	0.076	0.194	(-0.310, 0.452)
Drink	0.585	0.379	(-0.148, 1.327)	Age	-0.127	0.078	(-0.282, 0.021)
Age	-0.398	0.084	(-0.563, -0.237)	$\sum_{j=0}^{2} R_j$	1.776	0.103	(1.575, 1.976)
z				z			
Baseline	0.086	0.122	(-0.154, 0.328)	Baseline	-0.633	0.231	(-1.080, -0.173)
6-Month	-0.155	0.130	(-0.410, 0.100)	6-Month	-0.073	0.141	(-0.357, 0.198)
12-Month	-0.427	0.140	(-0.702, -0.158)	12-Month	0.456	0.142	(0.175, 0.736)
18-Month	-0.372	0.141	(-0.654, -0.105)	18-Month	0.133	0.148	(-0.149, 0.430)
ρ	0.792	0.036	(0.722, 0.860)	—	—	—	—
α	0.742	0.046	(0.652, 0.831)	—	—	—	—
τ	1.074	1.241	(0.000, 3.661)	—	—	—	—

Table 6: Posterior Summaries under the Nonignorable Model with a $N(0, 8)$ Prior for the HIV Prevention Behavioral Data

	Binary Response Model				Missing Data Mechanism		
	EST	SD	95% HPD Interval		EST	SD	95% HPD Interval
Baseline				**Baseline**			
Intercept	-0.678	0.193	(-1.062, -0.305)	Intercept	-3.632	0.740	(-4.870, -2.450)
Gender	0.375	0.129	(0.129, 0.639)	Gender	0.114	0.239	(-0.357, 0.578)
City	0.118	0.152	(-0.187, 0.409)	City	-0.329	0.325	(-0.986, 0.290)
Cohabit	0.702	0.139	(0.438, 0.980)	Cohabit	0.226	0.248	(-0.254, 0.719)
Counselor	0.422	0.157	(0.108, 0.724)	Counselor	0.655	0.369	(-0.056, 1.379)
Drink	0.416	0.345	(-0.252, 1.104)	Age	0.085	0.122	(-0.146, 0.333)
Age	-0.359	0.070	(-0.491, -0.217)	y_0	0.117	0.979	(-1.567, 1.934)
6-Month				**6-Month**			
Intercept	-1.630	0.288	(-2.225, -1.099)	Intercept	-2.209	0.332	(-2.820, -1.600)
Gender	0.111	0.142	(-0.180, 0.383)	Gender	-0.390	0.150	(-0.673, -0.083)
City	0.101	0.162	(-0.215, 0.415)	City	0.032	0.186	(-0.333, 0.396)
Cohabit	0.628	0.142	(0.344, 0.900)	Cohabit	0.190	0.160	(-0.127, 0.505)
Counselor	0.573	0.176	(0.226, 0.914)	Counselor	0.238	0.207	(-0.174, 0.634)
Drink	0.967	0.355	(0.301, 1.690)	Age	-0.069	0.095	(-0.248, 0.126)
Age	-0.451	0.081	(-0.606, -0.293)	R_0	0.344	0.313	(-0.278, 0.950)
—	—	—	—	y_0	-0.262	0.333	(-0.938, 0.347)
—	—	—	—	y_1	0.521	0.952	(-1.404, 2.367)
12-Month				**12-Month**			
Intercept	-1.501	0.304	(-2.093, -0.905)	Intercept	-2.331	0.385	(-3.060, -1.646)
Gender	0.216	0.152	(-0.069, 0.525)	Gender	-0.574	0.160	(-0.884, -0.255)
City	-0.037	0.170	(-0.369, 0.291)	City	-0.121	0.194	(-0.505, 0.255)
Cohabit	0.609	0.148	(0.318, 0.896)	Cohabit	-0.194	0.158	(-0.501, 0.117)
Counselor	0.263	0.178	(-0.080, 0.611)	Counselor	-0.260	0.187	(-0.615, 0.113)
Drink	0.518	0.356	(-0.177, 1.208)	Age	-0.100	0.089	(-0.272, 0.072)
Age	-0.493	0.087	(-0.667, -0.330)	$\sum_{j=0}^{1} R_j$	1.765	0.183	(1.408, 2.120)
—	—	—	—	y_1	-0.653	0.317	(-1.239, -0.015)
—	—	—	—	y_2	1.437	0.714	(0.035, 2.822)
18-Month				**18-Month**			
Intercept	-1.705	0.275	(-2.250, -1.192)	Intercept	-2.726	0.258	(-3.243, -2.234)
Gender	0.243	0.148	(-0.043, 0.535)	Gender	-0.403	0.156	(-0.699, -0.093)
City	-0.145	0.175	(-0.497, 0.196)	City	0.404	0.185	(0.046, 0.770)
Cohabit	0.472	0.144	(0.188, 0.752)	Cohabit	0.049	0.152	(-0.251, 0.344)
Counselor	0.387	0.179	(0.031, 0.736)	Counselor	0.087	0.197	(-0.296, 0.478)
Drink	0.569	0.364	(-0.127, 1.301)	Age	-0.107	0.082	(-0.269, 0.053)
Age	-0.386	0.082	(-0.551, -0.229)	$\sum_{j=0}^{2} R_j$	1.754	0.111	(1.532, 1.966)
—	—	—	—	y_2	0.604	0.291	(0.043, 1.169)
—	—	—	—	y_3	-0.4944	0.5608	(-1.562, 0.640)
z				z			
Baseline	0.084	0.119	(-0.147, 0.326)	Baseline	-0.637	0.233	(-1.111, -0.202)
6-Month	-0.158	0.127	(-0.410, 0.090)	6-Month	-0.049	0.149	(-0.349, 0.235)
12-Month	-0.372	0.140	(-0.646, -0.100)	12-Month	0.579	0.166	(0.269, 0.917)
18-Month	-0.357	0.137	(-0.631, -0.100)	18-Month	0.147	0.153	(-0.158, 0.443)
ρ	0.789	0.037	(0.716, 0.860)	—	—	—	—
α	0.727	0.048	(0.635, 0.825)	—	—	—	—
τ	1.117	1.280	(0.000, 3.825)	—	—	—	—

Table 7: Posterior Summaries under the Nonignorable Model with Jeffreys Prior for the HIV Prevention Behavioral Data

	Binary Response Model				Missing Data Mechanism		
	EST	SD	95% HPD Interval		EST	SD	95% HPD Interval
Baseline				**Baseline**			
Intercept	-0.675	0.195	(-1.059, -0.300)	Intercept	-3.559	0.639	(-4.880, -2.446)
Gender	0.373	0.130	(0.113, 0.623)	Gender	0.106	0.236	(-0.348, 0.568)
City	0.123	0.151	(-0.161, 0.431)	City	-0.301	0.319	(-0.939, 0.314)
Cohabit	0.704	0.141	(0.430, 0.973)	Cohabit	0.214	0.246	(-0.252, 0.711)
Counselor	0.422	0.155	(0.119, 0.723)	Counselor	0.608	0.355	(-0.074, 1.313)
Drink	0.430	0.345	(-0.236, 1.109)	Age	0.093	0.120	(-0.142, 0.327)
Age	-0.363	0.068	(-0.497, -0.230)	y_0	0.147	0.907	(-1.615, 2.037)
6-Month				**6-Month**			
Intercept	-1.650	0.287	(-2.223, -1.120)	Intercept	-2.147	0.280	(-2.690, -1.603)
Gender	0.117	0.142	(-0.169, 0.385)	Gender	-0.391	0.147	(-0.690, -0.114)
City	0.103	0.160	(-0.218, 0.406)	City	0.038	0.185	(-0.326, 0.393)
Cohabit	0.630	0.147	(0.353, 0.921)	Cohabit	0.191	0.159	(-0.117, 0.504)
Counselor	0.570	0.181	(0.210, 0.924)	Counselor	0.230	0.206	(-0.166, 0.641)
Drink	0.983	0.361	(0.277, 1.697)	Age	-0.073	0.092	(-0.250, 0.110)
Age	-0.454	0.079	(-0.610, -0.303)	R_0	0.333	0.311	(-0.288, 0.930)
—	—	—	—	y_0	-0.237	0.304	(-0.814, 0.363)
—	—	—	—	y_1	0.431	0.901	(-1.262, 2.011)
12-Month				**12-Month**			
Intercept	-1.546	0.313	(-2.172, -0.964)	Intercept	-2.243	0.313	(-2.864, -1.657)
Gender	0.232	0.153	(-0.066, 0.532)	Gender	-0.556	0.159	(-0.861, -0.235)
City	-0.030	0.173	(-0.362, 0.303)	City	-0.112	0.192	(-0.491, 0.260)
Cohabit	0.616	0.151	(0.337, 0.921)	Cohabit	-0.183	0.156	(-0.487, 0.123)
Counselor	0.268	0.182	(-0.091, 0.621)	Counselor	-0.263	0.186	(-0.621, 0.112)
Drink	0.541	0.363	(-0.175, 1.255)	Age	-0.103	0.087	(-0.270, 0.071)
Age	-0.500	0.085	(-0.672, -0.339)	$\sum_{j=0}^{1} R_j$	1.731	0.171	(1.399, 2.067)
—	—	—	—	y_1	-0.602	0.301	(-1.182, -0.002)
—	—	—	—,	y_2	1.2918	0.6383	(0.011, 2.532)
18-Month				**18-Month**			
Intercept	-1.732	0.288	(-2.288, -1.191)	Intercept	-2.688	0.252	(-3.190, -2.191)
Gender	0.248	0.151	(-0.046, 0.553)	Gender	-0.396	0.154	(-0.684, -0.082)
City	-0.140	0.182	(-0.494, 0.217)	City	0.408	0.184	(0.047, 0.766)
Cohabit	0.471	0.141	(0.194, 0.750)	Cohabit	0.055	0.152	(-0.233, 0.359)
Counselor	0.401	0.182	(0.054, 0.771)	Counselor	0.083	0.199	(-0.310, 0.475)
Drink	0.582	0.378	(-0.141, 1.352)	Age	-0.108	0.082	(-0.267, 0.052)
Age	-0.388	0.081	(-0.545, -0.228)	$\sum_{j=0}^{2} R_j$	1.741	0.109	(1.528, 1.955)
—	—	—	—	y_2	0.563	0.289	(-0.010, 1.111)
—	—	—	—	y_3	-0.473	0.550	(-1.554, 0.573)
z				z			
Baseline	0.084	0.120	(-0.149, 0.322)	Baseline	-0.623	0.227	(-1.085, -0.194)
6-Month	-0.155	0.125	(-0.406, 0.084)	6-Month	-0.052	0.147	(-0.336, 0.237)
12-Month	-0.379	0.136	(-0.657, -0.123)	12-Month	0.558	0.159	(0.245, 0.867)
18-Month	-0.357	0.140	(-0.641, -0.093)	18-Month	0.145	0.153	(-0.150, 0.446)
ρ	0.788	0.036	(0.718, 0.859)	—	—	—	—
α	0.731	0.047	(0.640, 0.826)	—	—	—	—
τ	1.059	1.211	(0.000, 3.567)	—	—	—	—

investigated in Wu *et al.* (2017), the posterior estimates under the Jeffreys prior using the all available data are similar to those under the Jeffreys prior using a subset of the data as long as the design matrix is of full rank. We expect that the posterior estimates are quite robust to the selection of the subset used in constructing Jeffreys prior.

We currently use the DIC ($\mathrm{DIC}_{\mathbf{R}|\mathbf{y}}$) and conditional LPML ($\mathrm{LPML}_{\mathbf{R}|\mathbf{y}}$) criteria to assess fit of the missing data mechanism. Our DIC ($\mathrm{DIC}_{\mathbf{R}|\mathbf{y}}$) is a part of the "conditional DIC" in Mason *et al.* (2012); Zhang *et al.* (2015), since the deviance function is defined based on the distribution of the missing data indicators conditional on the missing responses. Since our interest lies in the missing data mechanism, $\mathrm{DIC}_{\mathbf{R}|\mathbf{y}}$ may be more suitable in our application. As shown in Section 5, $\mathrm{DIC}_{\mathbf{R}|\mathbf{y}}$ has good empirical performance according to our simulation study. We also investigated the DIC and LPML of the joint model after integrating out the missing responses. However, the DIC and LPML of the joint model failed to assess the fit of the missing data mechanism in both the simulation study and the real data analysis. We note that similar results were also observed in Mason *et al.* (2012). Potential future research involves extending the current DIC and conditional LPML criteria to assess fit of the joint model via the decomposition of DIC and LPML (Zhang *et al.*, 2015). This future work is currently under investigation.

Acknowledgement

The authors would like to thank the Co-Guest Editor, the Associate Editor, and the two anonymous referees for their help comments and suggestions, which have lead to an improved version of the paper. Dr. Ibrahim and Dr. Chen's research was partially supported by NIH grants #GM70335 and #P01CA142538. Dr. Fisher's research was partially supported by NIH grant #5R01MH077524.

Appendix: Proofs

Proof of Proposition 4.1. If we assume $\pi(\boldsymbol{\gamma}) = 1$

$$\pi^*(\boldsymbol{\theta}|D_{\text{obs}}) = \mathcal{L}(\boldsymbol{\theta}|D_{\text{obs}})\pi(\boldsymbol{\beta}, \alpha, \tau, \rho)$$

$$= \sum_{y_{\text{mis}}} \prod_{i=1}^{n} \prod_{k=1}^{K} \left\{ \int f_y(\mathbf{y}_i|z_i, \mathbf{x}_{1i}, \boldsymbol{\epsilon}_i, \boldsymbol{\theta}) f(\boldsymbol{\epsilon}_i|\alpha, \rho) d\boldsymbol{\epsilon} f(\zeta_k|\tau) d\boldsymbol{\zeta} f_{\mathbf{R}|\mathbf{y}}(\mathbf{R}_{iT}|\mathbf{y}_i, z_i, \mathbf{x}_{2i}, \boldsymbol{\gamma}_t)\pi(\boldsymbol{\beta}, \alpha, \rho) \right\}.$$

Define $y_{it}^* = y_{it}$ if $r_{it} = 0$, and $y_{it}^* = 0$ if $r_{it} = 1$. Let $\mathbf{y}_i^* = (y_{i0}^*, \ldots, y_{iT}^*)$. It can be shown that

$$\pi^*(\boldsymbol{\theta}|D_{\text{obs}}) \geq \prod_{i=1}^{n}\prod_{k=1}^{K}\Big\{\int f_y(\mathbf{y}_i^*|z_i, \mathbf{x}_{1i}, \boldsymbol{\epsilon}_i, \boldsymbol{\theta}) f(\boldsymbol{\epsilon}_i|\alpha, \rho) d\boldsymbol{\epsilon} f(\zeta_k|\tau) d\boldsymbol{\zeta}$$
$$\prod_{t=0}^{T} f_{\mathbf{R}|\mathbf{y}}(R_{it}|\mathbf{R}_{it-1}, \mathbf{y}_i^*, z_i, \mathbf{x}_{2i}, \boldsymbol{\gamma}_t)\pi(\boldsymbol{\beta}, \alpha, \tau, \rho)\Big\}.$$

Note that for each t, the unnormalized marginal posterior density of $\boldsymbol{\gamma}_t$ with $\pi(\boldsymbol{\gamma}_t) = 1$ is $\prod_{i=1}^n f(R_{it}|\mathbf{R}_{it-1}, \mathbf{y}_i^*, z_i, \mathbf{x}_{2i}, \boldsymbol{\gamma}_t)$, which corresponds to a binary regression model with response equal to R_{it}. Due to the construction of y_i^* and Proposition A.1 (Huang *et al.*, 2005), the posterior density of $\boldsymbol{\gamma}_t$ is improper and thus the joint posterior $\pi^*(\boldsymbol{\theta}|D_{\text{obs}})$ is also improper.

Proof of Proposition 4.2. Because $f_{\mathbf{R}|\mathbf{y}}(\mathbf{R}_{iT}|\mathbf{y}_i, z_i, \mathbf{x}_{2i}, \boldsymbol{\gamma}_t) \leq 1$, $\pi(\boldsymbol{\gamma})$ and $\pi(\tau)$ are proper, and we assume $\pi(\boldsymbol{\beta}, \boldsymbol{\alpha}, \rho) = 1$, it suffices to show that

$$\int \sum_{y_{\text{mis}}} \prod_{i=1}^{n}\prod_{k=1}^{K} \int f_y(\mathbf{y}_i|z_i, \mathbf{x}_{1i}, \boldsymbol{\epsilon}_i, \boldsymbol{\theta}) f(\boldsymbol{\epsilon}_i|\alpha, \rho) d\boldsymbol{\epsilon} f(\zeta_k|\tau) d\boldsymbol{\zeta} \pi(\tau) d\tau d\boldsymbol{\beta} d\boldsymbol{\alpha} d\rho < \infty. \qquad (7.1)$$

Let $\mathbf{y}^* = (\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}^*)$, where $\mathbf{y}_{\text{mis}}^*$ is any combination of the possible values for the missing responses. Due to the finite number of combinations of $\mathbf{y}_{mis}^*$ and by Tonelli's theorem, it suffices to show that for each k

$$\prod_{i \in I_c} \int f_y(\mathbf{y}_i^*|z_i, \mathbf{x}_{1i}, \boldsymbol{\epsilon}_i, \boldsymbol{\theta}) d\boldsymbol{\beta} f(\boldsymbol{\epsilon}_i|\alpha, \rho) d\boldsymbol{\epsilon} f(\zeta_k|\tau) d\boldsymbol{\zeta} \pi(\tau) d\tau d\boldsymbol{\alpha} d\rho < \infty.$$

By Chen and Shao (2001), and under (C1) and (C2), there exists a constant K_0 depending only on $\mathbf{X}_{obs}^*$ such that

$$\begin{aligned}
&\prod_{i \in I_c} \int f_y(\mathbf{y}_i^*|z_i, \mathbf{x}_{1i}, \boldsymbol{\epsilon}_i, \boldsymbol{\theta}) d\boldsymbol{\beta} f(\boldsymbol{\epsilon}_i|\alpha, \rho) d\boldsymbol{\epsilon} f(\zeta_k|\tau) d\boldsymbol{\zeta} \pi(\tau) d\tau d\boldsymbol{\alpha} d\rho \\
=&E_{\mathbf{u}}\left(\int \mathbf{1}(\mathbf{X}_{\text{obs}}^*\boldsymbol{\beta} + \tau\boldsymbol{\zeta} + \boldsymbol{\epsilon} \leq \mathbf{u}) d\boldsymbol{\beta} f(\boldsymbol{\epsilon}|\alpha, \rho) d\boldsymbol{\epsilon} f(\zeta_k|\tau) d\boldsymbol{\zeta} \pi(\tau) d\tau d\boldsymbol{\alpha} d\rho\right) \\
=&E_{\mathbf{u}}\left(\int K_0\|\mathbf{u} - \tau\boldsymbol{\zeta} - \boldsymbol{\epsilon}\|^p d\boldsymbol{\beta} f(\boldsymbol{\epsilon}|\alpha, \rho) d\boldsymbol{\epsilon} f(\zeta_k|\tau) d\boldsymbol{\zeta} \pi(\tau) d\tau d\boldsymbol{\alpha} d\rho\right) \\
\leq&E_{\mathbf{u}}\left(K_0\|\mathbf{u}\|^p \int f(\boldsymbol{\epsilon}|\alpha, \rho) d\boldsymbol{\epsilon} f(\zeta_k|\tau) d\boldsymbol{\zeta} \pi(\tau) d\tau d\boldsymbol{\alpha} d\rho\right) + \\
&K_0 \int \|\boldsymbol{\zeta}\|^p f(\zeta_k|\tau) d\boldsymbol{\zeta} \tau^p \pi(\tau) d\tau f(\boldsymbol{\epsilon}|\alpha, \rho) d\boldsymbol{\epsilon} d\boldsymbol{\alpha} d\rho + \\
&K_0 \int \|\boldsymbol{\epsilon}\|^p f(\boldsymbol{\epsilon}|\alpha, \rho) d\boldsymbol{\epsilon} f(\zeta_k|\tau) d\boldsymbol{\zeta} \pi(\tau) d\tau d\boldsymbol{\alpha} d\rho.
\end{aligned}$$

The first term and second term are finite since $\boldsymbol{\alpha} \in (0,1)$, $\rho \in (-1,1)$, $\pi(\tau)$ is proper with a finite p^{th} moment, $\zeta_k \overset{i.i.d.}{\sim} N(0,1)$, and condition C3. Let $\Sigma = \Gamma\Gamma$, where $\Gamma = \Gamma'$. To study the second term, we first carry out a transformation on $\boldsymbol{\epsilon}_i$ such that $\boldsymbol{\epsilon}_i^* = (\sqrt{\alpha}\Gamma)^{-1}\boldsymbol{\epsilon}_i, \quad i \in I_c$. Write the second term as

$$
\begin{aligned}
&K_0 \int \|\boldsymbol{\epsilon}\|^p f(\boldsymbol{\epsilon}|\alpha,\rho)d\boldsymbol{\epsilon} f(\zeta_k|\tau)d\boldsymbol{\zeta}\pi(\tau)d\tau d\boldsymbol{\alpha} d\rho \\
\le& K_0 \int \sum_{i \in I_c} \|\boldsymbol{\epsilon}_i\|^p f(\boldsymbol{\epsilon}|\alpha,\rho)d\boldsymbol{\epsilon} f(\zeta_k|\tau)d\boldsymbol{\zeta}\pi(\tau)d\tau d\boldsymbol{\alpha} d\rho \\
=& K_0 \sum_{i \in I_c} \int \|\boldsymbol{\epsilon}_i\|^p f(\boldsymbol{\epsilon}|\alpha,\rho)d\boldsymbol{\epsilon} f(\zeta_k|\tau)d\boldsymbol{\zeta}\pi(\tau)d\tau d\boldsymbol{\alpha} d\rho \\
=& K_0 \sum_{i \in I_c} \int \|\boldsymbol{\epsilon}_i\|^p f(\boldsymbol{\epsilon}_i|\alpha,\rho)d\boldsymbol{\epsilon}_i f(\zeta_k|\tau)d\boldsymbol{\zeta}\pi(\tau)d\tau d\boldsymbol{\alpha} d\rho \\
=& \frac{K_0}{\sqrt{2\pi}} \sum_{i \in I_c} \int \|\boldsymbol{\epsilon}_i\|^p \frac{1}{|\alpha\Sigma|^{1/2}} \exp\left(-\frac{\boldsymbol{\epsilon}_i'\Sigma^{-1}\boldsymbol{\epsilon}_i}{2\alpha}\right) d\boldsymbol{\epsilon}_i f(\zeta_k|\tau)d\boldsymbol{\zeta}\pi(\tau)d\tau d\boldsymbol{\alpha} d\rho \\
=& \frac{K_0}{\sqrt{2\pi}} \sum_{i \in I_c} \int \left(\boldsymbol{\epsilon}_i^{*\prime}\alpha\Sigma\boldsymbol{\epsilon}_i^*\right)^{p/2} \exp\left(-\frac{\|\boldsymbol{\epsilon}_i^*\|^2}{2}\right) d\boldsymbol{\epsilon}_i^* f(\zeta_k|\tau)d\boldsymbol{\zeta}\pi(\tau)d\tau d\boldsymbol{\alpha} d\rho.
\end{aligned}
$$

Let λ_{max} denote the maximum eigenvalues of Σ, and when $T+1=4$, $\lambda_{max} < 4$ given $\rho \in (-1,1)$. We also know that $\boldsymbol{\epsilon}_i^{*\prime}\Sigma\boldsymbol{\epsilon}_i^* \le \lambda_{max}\|\boldsymbol{\epsilon}_i^*\|^2$. Therefore,

$$
\begin{aligned}
LHS \le& \frac{K}{\sqrt{2\pi}} \sum_{i \in I_c} \int \alpha^{p/2} \left\{4\|\boldsymbol{\epsilon}_i^*\|^2\right\}^{p/2} \exp\left(-\frac{\|\boldsymbol{\epsilon}_i^*\|^2}{2}\right) d\boldsymbol{\epsilon}_i^* f(\zeta_k|\tau)d\boldsymbol{\zeta}\pi(\tau)d\tau d\boldsymbol{\alpha} d\rho \\
\le& K' \sum_{i \in I_c} \sum_{t=0}^{T} \int \alpha^{p/2} |\epsilon_{it}^*|^p \exp\left(-\frac{\epsilon_{it}^{*\,2}}{2}\right) d\epsilon_{it}^* f(\zeta_k|\tau)d\boldsymbol{\zeta}\pi(\tau)d\tau d\boldsymbol{\alpha} d\rho,
\end{aligned}
$$

where K' is some constant depending only on $\mathbf{X}^*_{\text{obs}}$. Again, since $\alpha \in (0,1)$, $\rho \in (-1,1)$, $\pi(\tau)$ is propoer, and $\zeta_k \overset{i.i.d.}{\sim} N(0,1)$, the second term is also finite, which together yields (7.1).

References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, **88**(422), 669–679.

Albert, P. S. (2000). A transitional model for longitudinal binary data subject to nonignorable missing data. *Biometrics*, **56**(2), 602–608.

Chen, M.-H. and Shao, Q.-M. (2001). Propriety of posterior distribution for dichotomous quantal response models. *Proceedings of the American Mathematical Society*, **129**(1), 293–302.

Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2000). *Monte Carlo methods in Bayesian computation*. New York: Springer.

Chen, M.-H., Ibrahim, J. G., and Kim, S. (2008). Properties and implementation of Jeffreys's prior in binomial regression models. *Journal of the American Statistical Association*, **103**(484), 1659–1664.

Chib, S. and Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, **85**(2), 347–361.

Fisher, J. D., Cornman, D. H., Shuper, P. A., Christie, S., Pillay, S., Macdonald, S., Ngcobo, N., Amico, K. R., Lalloo, U., Friedland, G., *et al.* (2014). HIV prevention counseling intervention delivered during routine clinical care reduces HIV risk behavior in HIV-infected South Africans receiving antiretroviral therapy: the Izindlela Zokuphila/Options for health randomized trial. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, **67**(5), 499–507.

Follmann, D. and Wu, M. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics*, **51**, 151–168.

Glynn, R. J., Laird, N. M., and Rubin, D. B. (1986). Selection modeling versus mixture modeling with nonignorable nonresponse. In *Drawing inferences from self-selected samples*, pages 115–142. Springer.

Hanson, T. E., Branscum, A. J., and Johnson, W. O. (2011). Predictive comparison of joint longitudinal-survival modeling: a case study illustrating competing approaches. *Lifetime Data Analysis*, **17**(1), 3–28.

Huang, L., Chen, M.-H., and Ibrahim, J. G. (2005). Bayesian analysis for generalized linear models with nonignorably missing covariates. *Biometrics*, **61**(3), 767–780.

Ibrahim, J. G., Lipsitz, S. R., and Chen, M.-H. (1999). Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **61**(1), 173–190.

Ibrahim, J. G., Chen, M.-H., and Lipsitz, S. R. (2001). Missing responses in generalised linear mixed models when the missing data mechanism is nonignorable. *Biometrika*, **88**(2), 551–564.

Ibrahim, J. G., Chen, M.-H., Lipsitz, S. R., and Herring, A. H. (2005). Missing-data methods for generalized linear models: A comparative review. *Journal of the American Statistical Association*, **100**(469), 332–346.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 186 (1007), pages 453–461. The Royal Society.

Kim, S., Chen, M.-H., and Dey, D. K. (2008). Flexible generalized t-link models for binary response data. *Biometrika*, **95**(1), 93–106.

Little, R. J. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, **88**(421), 125–134.

Little, R. J. and Rubin, D. B. (2002). *Statistical analysis with missing data*. New York: John Wiley & Sons, 1 edition.

Liu, J. S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, **89**(427), 958–966.

Mason, A., Richardson, S., and Best, N. (2012). Two-pronged strategy for using DIC to compare selection models with non-ignorable missing responses. *Bayesian Analysis*, **7**(1), 109–146.

Newman, D. A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods*, **6**(3), 328–362.

Preisser, J. S., Lohman, K. K., and Rathouz, P. J. (2002). Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random. *Statistics in Medicine*, **21**(20), 3035–3054.

Robins, J. M. and Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, **90**(429), 122–129.

Royston, P. and others (2004). Multiple imputation of missing values. *Stata Journal*, **4**(3), 227–41.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, **63**(3), 581–592.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*, volume 81. New York: John Wiley & Sons.

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., and Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj*, **338**, b2393.

Verbeke, G. (2005). *Models for Discrete Longitudinal Data. Springer Series in Statistics*. New York: Springer.

Wu, J., de Castro, M., Schifano, E. D., and Chen, M.-H. (2017). Assessing covariate effects using Jeffreys-type prior in the Cox model in the presence of a monotone partial likelihood. *Journal of Statistical Theory and Practice*, **In press**.

Yao, H., Kim, S., Chen, M.-H., Ibrahim, J. G., Shah, A. K., and Lin, J. (2015). Bayesian inference for multivariate meta-regression with a partially observed within-study sample covariance matrix. *Journal of the American Statistical Association*, **110**(510), 528–544.

Zhang, D., Chen, M.-H., Ibrahim, J. G., Boye, M. E., Wang, P., and Shen, W. (2014). Assessing model fit in joint models of longitudinal and survival data with applications to cancer clinical trials. *Statistics in Medicine*, **33**(27), 4715–4733.

Zhang, J., Chu, H., Hong, H., Virnig, B. A., and Carlin, B. P. (2015). Bayesian hierarchical models for network meta-analysis incorporating nonignorable missingness. *Statistical Methods in Medical Research*, **doi: 10.1177/0962280215596185**.