5-25-2017

Bayesian Modeling and Inference for Nonignorably Missing Longitudinal Response Data

Jing Wu
jing.wu@uconn.edu

Bayesian Modeling and Inference for

Nonignorably Missing Longitudinal Response Data

Jing Wu, Ph.D.

University of Connecticut, 2017

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 dissertation research, we develop models and carry out Bayesian inferences for both longitudinal binary response and count response data. For longitudinal binary response data, we develop a new probit model. It resolves the well-known weak identifiability issue of the variance of the random effects, and substantially improves the convergence and mixing of the Gibbs sampling algorithm. We adopt a sequence of one-dimensional conditional distributions for the missing data indicators via a logistic regression model. For the longitudinal count response data, we use the zero-inflated Poisson model for the response measurements, and propose a new conditional model for the missing data mechanism. The new model has the potential of reducing the number of nuisance parameters, allows us to model dropout and intermittent missing jointly, and provides a broad class of missing data mechanisms. We then investigate and characterize the conditions for propriety of the joint posterior distribution under both binary and Poisson cases, and propose a variation of Jeffreys's prior as a remedy for impropriety of the posterior. In addition, we develop two efficient Gibbs sampling algorithms for both binary and Poisson cases, which allow us to conveniently sample missing responses and to apply the collapsed Gibbs technique as well as the hierarchical centering technique within the Gibbs sampling framework.

The proposed methodologies and the sampling techniques are illustrated using real data from an HIV prevention clinical trial. A sensitivity analysis is carried out to assess the robustness of the posterior estimates under different prior specifications and missing data mechanisms. Two model assessment criteria, the deviance information criterion (DIC) and the logarithm of the pseudomarginal likelihood (LPML), are used to examine model fit. Extensive real data analyses are conducted to assess the performances of missing data mechanisms under different scenarios.

Bayesian Modeling and Inference for

Nonignorably Missing Longitudinal Response Data

Jing Wu

B.S., Shanghai Jiao Tong University, China, 2012

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2017

APPROVAL PAGE

Doctor of Philosophy Dissertation

Bayesian Modeling and Inference for

Nonignorably Missing Longitudinal Response Data

Presented by

Jing Wu, B.S.

Major Advisor ______________________________
Ming-Hui Chen

Major Advisor ______________________________
Elizabeth D. Schifano

Associate Advisor ______________________________
Jun Yan

University of Connecticut

2017

To my mother Yalian Chen and my father Guohui Wu

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deep gratitude to my major advisor Prof. Ming-Hui Chen for his continuous support through all these years. Prof. Chen has supported me not only by providing the unique assistantship and collaboration opportunities, but also being an incredibly wonderful academic and life mentor. I would like to thank him for leading me to the research area of statistics, partnering with me on the accumulation of advanced statistical knowledge, and inspiring me to pursue teaching and research as my lifetime career. Choosing him as my advisor is one of the best decisions I have made.

I gratefully acknowledge my co-major advisor Prof. Elizabeth D. Schifano for her great guidance and insightful advice, which are extremely important to me in the whole process of my graduate study.

I would like to thank my associate advisor Prof. Jun Yan for all the kind help and advice on my job application and research work.

Special thanks to Pfizer Inc. for offering me the research fellow position and providing me the precious two-year industry experience. I would like to especially thank Dr. Huaming Tan and Dr. Neal Thomas for imparting me the pharmaceutical knowledge, and giving me advice on both professional and personal development.

My sincere thanks also goes to the Center for Nursing Research for supporting my half-time graduate assistantship during the past three academic years. I am particularly thankful to Ms. Elise Bennett, Prof. Xiaomei Cong, Prof. Stephen Walsh, and Wanli Xu for their warmth and support.

It is my great honor to work with and learn from Dr. Anthony V. D'Amico of Harvard Medical School. This invaluable opportunity substantially enhances my understanding of survival analysis.

Many thanks are owed to my fellow graduate students. It is their precious friendship and company that makes my journey to the Ph.D. more delightful. I would also like to thank Ms. Tracy Burke and Ms. Megan Petsa for their helpful administrative assistance.

Lastly, and most importantly, I want to dedicate this dissertation to my mother Yalian Chen and my father Guohui Wu, who always encourage and support me to pursue the life and career I adore. It is your unconditional love that leads me to conquer and to achieve.

TABLE OF CONTENTS

Chapter 1: Introduction **1**

1.1 Literature Review 1

1.2 HIV Prevention Behavioral Intervention Clinical Trials 3

1.3 Motivating Data 5

1.4 Methodologies Overview 9

1.5 Dissertation Outline 11

Chapter 2: Models for Longitudinal Binary Response Data **12**

2.1 The Proposed Models 12

2.1.1 The Model for Longitudinal Binary Measurements 12

2.1.2 Missing Data Mechanism 15

2.2 Bayesian Inference 16

2.2.1 The Likelihood Function 16

2.2.2 Prior and Posterior Distributions 17

2.2.3 Computational Development 20

2.2.4 Bayesian Model Assessment 25

2.3 A Simulation Study 27

2.4 Analysis of the HIV Prevention Behavioral Data 32

Chapter 3: Models for Longitudinal Count Response Data **42**

3.1 The Proposed Methods 42

3.1.1 The Model for Longitudinal Counts 42

3.1.2 Prequential Multinomial Model for Dropout and Conditional Model for Intermittent Missing 44
3.1.3 Predictive Probabilities of Dropout and Intermittent Missing 48
3.2 Bayesian Inference 49
3.2.1 The Likelihood Function 49
3.2.2 Adjusted Intervention Effect 50
3.2.3 Prior and Posterior Distributions 51
3.2.4 Computational Development 53
3.2.5 Bayesian Model Assessment 60
3.3 Analysis of the HIV Prevention Behavioral Data 62

Chapter 4: Assessment of Missing Data Mechanism 73

4.1 Motivation 73
4.2 Assessment of Missing Data Mechanism in Section 2.1.2 74
4.3 Assessment of Missing Data Mechanism in Section 3.1.2 77

Chapter 5: Extension and Future Research Direction 87

5.1 Extensions to Choice of Link Functions 87
5.2 Extensions to Missing Data Mechanism and Multiple Longitudinal Outcomes 88
5.2.1 Path Specific Model 88
5.2.2 Joint Modeling of Multiple Types of Responses 90
5.3 Extensions to Model Selection Criteria 91

Appendix A: Proofs of Theorems 92

Appendix B: Additional Tables 97

Bibliography **100**

LIST OF TABLES

1.1 Characteristics of Study Participants (N=1875) 6

1.2 Missing Pattern of the Count Responses (ACASI-reported number of unprotected sex acts). 8

2.1 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% 33

2.2 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% 34

2.3 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 35

2.4 Posterior Summaries under the Ignorable Model for the HIV Prevention Behavioral Data 39

2.5 Posterior Summaries under the Nonignorable Model with a $N(0, 8)$ Prior for the HIV Prevention Behavioral Data 40

2.6 Posterior Summaries under the Nonignorable Model with Jeffreys Prior for the HIV Prevention Behavioral Data 41

3.1 Values of $\text{DIC}_{W,\mathbf{R}|\mathbf{y}}$ (p_D) and $\text{LPML}_{W,\mathbf{R}|\mathbf{y}}$ under Ignorable Missingness and Nonignorable Missingness with Various Priors for the HIV Prevention Behavioral Data 63

3.2 Posterior Summaries under the Ignorable Model for the HIV Prevention Behavioral Data. 68
3.3 Posterior Summaries under the Nonignorable Model with a $N(0,9)$ Prior for the HIV Prevention Behavioral Data. 69
3.4 Posterior Summaries under the Nonignorable Model with Jeffreys Prior for the HIV Prevention Behavioral Data. 70
3.5 Posterior Summaries of the Adjusted Overall Intervention Effect under the Ignorable Model, Nonignorable Model with a N(0, 9) Prior and Jeffreys Prior. 72
4.1 Values of $\text{DIC}_{\mathbf{R}|\mathbf{y}}$ (p_D) and $\text{LPML}_{\mathbf{R}|\mathbf{y}}$ for Longitudinal Binary Variable and Longitudinal Poisson Variable under Ignorable Missingness and Nonignorable Missingness with Various Priors for the HIV Prevention Behavioral Data 75
4.2 Posterior Summaries of the Adjusted Overall Intervention Effect under the Ignorable Model, Nonignorable Model with a N(0, 7) Prior and Jeffreys Prior. 77
4.3 Posterior Summaries under the Ignorable Model for the HIV Prevention Behavioral Data. 78
4.4 Posterior Summaries under the Nonignorable Model with a $N(0,7)$ Prior for the HIV Prevention Behavioral Data. 79
4.5 Posterior Summaries under the Nonignorable Model with Jeffreys Prior for the HIV Prevention Behavioral Data. 80
4.6 Values of $\text{DIC}_{W,\mathbf{R}|\mathbf{y}}$ (p_D) and $\text{LPML}_{W,\mathbf{R}|\mathbf{y}}$ for Longitudinal Binary Variable and Longitudinal Poisson Variable under Ignorable Missingness and Nonignorable Missingness with Various Priors for the HIV Prevention Behavioral Data 81

4.7 Posterior Summaries under the Ignorable Model for the HIV Prevention Behavioral Data. 83

4.8 Posterior Summaries under the Nonignorable Model with a $N(0, 8)$ Prior for the HIV Prevention Behavioral Data. 84

4.9 Posterior Summaries under the Nonignorable Model with Jeffreys Prior for the HIV Prevention Behavioral Data. 85

B.1 Posterior Summaries under the Nonignorable Model with a N(0, 1) Prior and a N(0, 10) Prior when missing percentage is low. 97

B.2 Posterior Summaries under the Nonignorable Model with a N(0, 1) Prior and a N(0, 10) Prior when missing percentage is high. 98

LIST OF FIGURES

1.1 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. 7

1.2 Trellis plots of the count responses (ACASI-reported number of unprotected sex acts). 9

2.1 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%. 30

Chapter 1

Introduction

1.1 Literature Review

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.

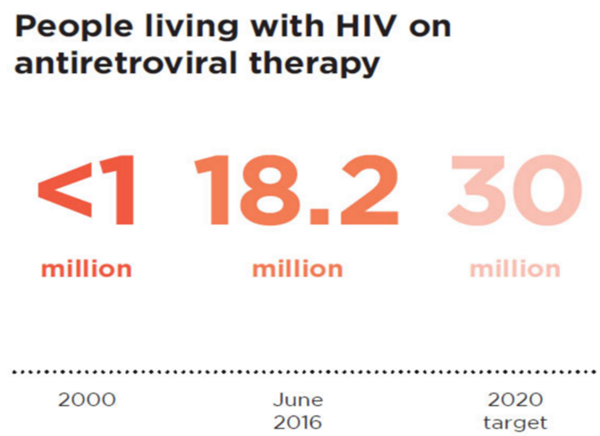
One approach for handling missing data is listwise deletion, in which all cases with missing values are deleted. This approach, however, introduces bias if the missingness is not MCAR. For MAR, inferential methods include maximum likelihood (Rubin, 1976; Ibrahim *et al.*, 1999; Newman, 2003; Ibrahim *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).

1.2 HIV Prevention Behavioral Intervention Clinical Trials

HIV, the human immunodeficiency virus, is a virus that attacks cells of the immune system (CD4) and interferes with the body's ability to fight infections. If left untreated, HIV will ultimately lead to acquired immunodeficiency syndrome (AIDS). According to the statistics from AIDS.GOV, 36.7 (0.5%) million people worldwide are currently suffering from HIV/AIDS. So far, there is no treatment that can eradicate the HIV virus. The most effective therapy against HIV is called antiretroviral therapy (ART), which is the combination of several antiretroviral medicines used to suppress the progression of HIV disease.

In addition to the medical treatment, HIV prevention behavioral intervention also plays a critical role in reducing the unprotected sexual risk behavior, and prevent the growth of the virus. As has been widely recognized, HIV treatment as prevention should be bundled with behavioral interventions to maximize effectiveness (Kalichman *et al.*, 2011).



More and more emphasis has been placed on HIV prevention behavioral research. Koblin *et al.* (2012) tested the efficacy of an HIV prevention behavioral intervention to reduce sexual risk among African-American men who have sex with men (MSM). Kalichman *et al.* (2011) concluded that a theory-based integrated behavioral intervention can improve HIV treatment adherence and reduce HIV transmission risks. Moreover, there is a need for combination prevention as there is for combination treatment. Combination prevention should be based on scientifically derived evidence, with input and engagement from local communities that fosters the successful integration of care and treatment (Bekker *et al.*, 2012).

In this dissertation research, we consider the data from an HIV prevention behavioral intervention clinical trial (Fisher *et al.*, 2014) in South Africa from June 2008 to May 2010, where people living with HIV (PLWH) on antiretroviral therapy (ART) constitute a large population. However, a significant proportion of them do not achieve viral suppression. They serve as relatively healthy but infectious vectors for transmission of HIV virus. PLWH who engage in unprotected sex also place themselves at risk for other sexually transmitted infections, associated morbidity, and accelerated progression of HIV disease.

To reduce the risk, an one-on-one counseling session with trained lay counselors concerning sexual risk behavior reduction is introduced. 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.

1.3 Motivating Data

The data from the HIV prevention behavioral intervention clinical trial we consider 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).

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

Table 1.1: 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.

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.

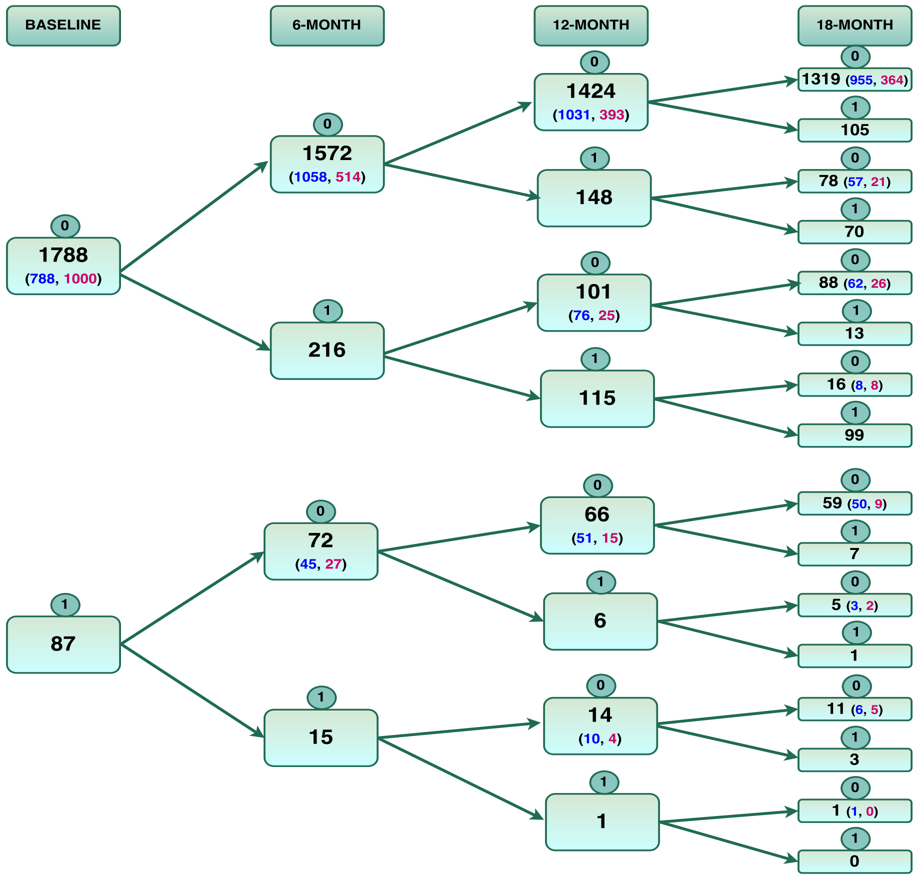


Figure 1.1: Path Diagram of the binary responses (any unprotected sex acts), where 0 in circle indicates observed and 1 in circle indicates missing; and the two numbers in parentheses indicate the number of zero counts (the first, blue) and the number of ones (the second, red) of the binary response 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 1.1 shows the characteristics of these 1875 PLWH, and Figure 1.1 visually presents the path diagram of the longitudinal binary response data (any unprotected sex acts).

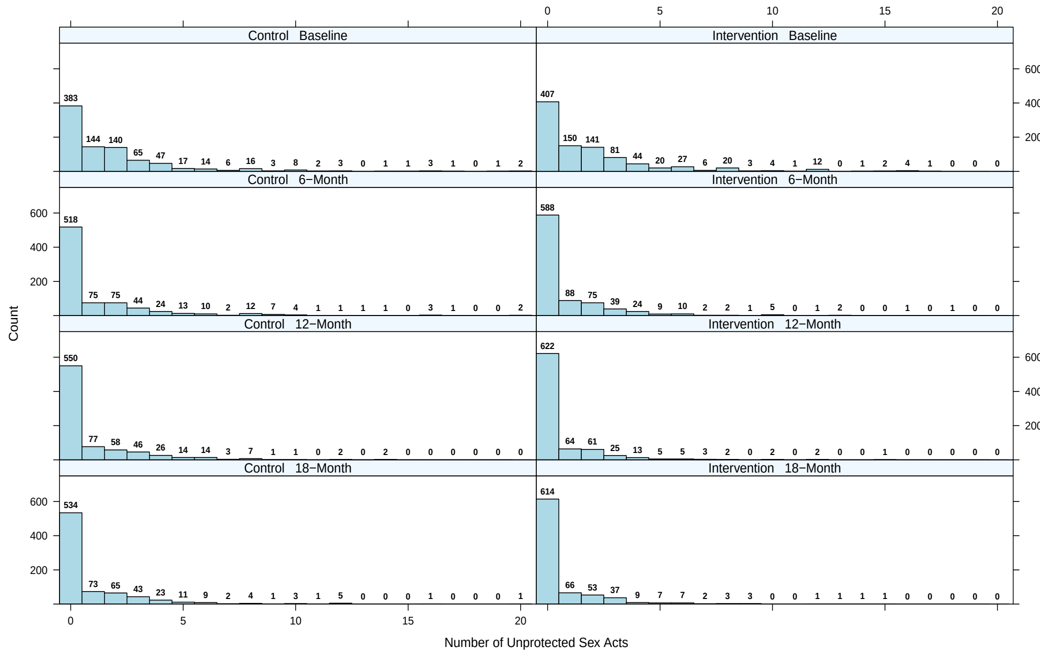
The longitudinal Poisson response variable we considered is the total number of ACASI-reported unprotected penile-vaginal or penile-anal sex acts in the past 4 weeks with partners of any HIV status. Therefore, the longitudinal Poisson response variable and the binary response variable are highly correlated. Furthermore, the two types of longitudinal measurements share with the same missing data pattern. We also expect that the longitudinal Poisson response variable contains more information than than the longitudinal binary response variable. Table 2.3 summarizes the missing pattern of the longitudinal count response data (ACASI-reported number of unprotected sex acts) and the number of unprotected sex acts by visit and treatment are visually exhibited in Figure 1.2.

Determining whether missing responses are ignorable or nonignorable is of great practical interest in HIV intervention clinical trials, which greatly motivates our proposed methodology.

Table 1.2: Missing Pattern of the Count Responses (ACASI-reported number of unprotected sex acts).

Pattern\Condition	Standard of Care	Intervention	Total
Completely Observed	646	673	1319
Intermittent Missing Only	130	128	258
Dropout w/ Intermittent Missing	10	14	24
Dropout w/o Intermittent Missing	129	145	274
Dropout at Baseline	4	3	7
Total	919	963	1882(1875)

Figure 1.2: Trellis plots of the count responses (ACASI-reported number of unprotected sex acts).



1.4 Methodologies Overview

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 Chapter 2, 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.

For the longitudinal count response data in Chapter 3, we apply the zero-inflated Poisson model for the response measurements, and propose a new conditional model for the missing data mechanism. The new model has the potential of reducing the number of nuisance parameters, allows us to model dropout and intermittent missing jointly, and provides a broad class of missing data mechanisms, which includes the sequential conditional model (Ibrahim *et al.*, 2001, 2005) as one special case. We then investigate and characterize the conditions for propriety of the joint posterior distribution under the new models, and propose a variation of Jeffreys prior as a remedy for impropriety of the posterior. In addition, we develop an efficient Gibbs sampling algorithm, which allows us to conveniently sample missing responses and to apply the hierarchical centering technique (Chen *et al.*, 2000) within the Gibbs sampling framework.

1.5 Dissertation Outline

The rest of the thesis is organized as follows. In Chapter 2, we introduce a new probit model with latent variables, and presents a joint multinomial model for the missing data indicators. We then 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 provide a detailed formulation of the partial DIC and conditional LPML criteria in the presence of missing data. An extensive simulation and a detailed analysis of the HIV prevention behavioral data are carried out in the end of this chapter.

Chapter 3 presents the model for longitudinal Poisson response variable, as well as the new missing data mechanisms for dropouts and intermittent missing. Again, we investigate and characterize the conditions for propriety of the joint posterior distribution, and provide a variation of Jeffreys prior to resolve the improper issue of the posterior. We then conduct the HIV prevention behavioral data analysis, using a new efficient Gibbs sampling algorithm. Extensive real data analyses are conducted in Chapter 4 to assess the performances of missing data mechanism under different scenarios. Future research directions are given in Chapter 5. The proofs of theorems are given in Appendix A. The additional tables are given in Appendix B.

Chapter 2

Models for Longitudinal Binary Response Data

2.1 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.

2.1.1 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

$$P(y_t = 1|z, \mathbf{x}_1, k, \boldsymbol{\beta}^*, \tau^*, \zeta_k, \epsilon_t^*) = \Phi(z\beta_{1t}^* + \mathbf{x}_1'\boldsymbol{\beta}_{2t}^* + \tau^*\zeta_k + \epsilon_t^*), \tag{2.1}$$

for $t = 0, \ldots, T$, where Φ is the $N(0,1)$ cumulative distribution function, $\mathbf{x}_1$ is a vector of baseline covariates, $\boldsymbol{\beta}^* = (\beta^*_{1t}, \boldsymbol{\beta}^{*\prime}_{2t})'$ 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 (2.1) by introducing the latent variables $\mathbf{w}^* = (w^*_0, \ldots, w^*_T)$. Following Albert and Chib (1993), (2.1) can be reformulated as

$$y_t = \begin{cases} 1 & \text{if } w^*_t \geq 0, \\ 0 & \text{if } w^*_t < 0, \end{cases} \tag{2.2}$$

and

$$w^*_t \mid \epsilon^*_t \sim N(z\boldsymbol{\beta}^*_{1t} + \mathbf{x}'_1\boldsymbol{\beta}^*_{2t} + \tau^*\zeta_k + \epsilon^*_t, 1) \tag{2.3}$$

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 (2.2) is invariant with respect to the scale parameter (variance) of w^*_t. To be more specific, if we replace w^*_t in (2.3) by $C \cdot w^*_t$, where C is any nonnegative constant, (2.2) is still identical to (2.1). 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 2.1). 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:

$$w_t = \frac{w_t^*}{\sqrt{1+\sigma^2}}, \quad \boldsymbol{\beta}_t = \frac{\boldsymbol{\beta}_t^*}{\sqrt{1+\sigma^2}}, \quad \tau = \frac{\tau^*}{\sqrt{1+\sigma^2}}, \quad \epsilon_t = \frac{\epsilon_t^*}{\sqrt{1+\sigma^2}}. \tag{2.4}$$

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

$$P(y_t = 1|z, \mathbf{x}_1, k, \boldsymbol{\beta}, \tau, \zeta_k, \epsilon_t) = \Phi\{(z\beta_{1t} + \mathbf{x}_1'\boldsymbol{\beta}_{2t} + \tau\zeta_k + \epsilon_t)\sqrt{1+\sigma^2}\} = \pi_t, \tag{2.5}$$

or

$$y_t = \begin{cases} 1 & \text{if } w_t \geq 0, \\ 0 & \text{if } w_t < 0, \end{cases} \tag{2.6}$$

and

$$w_t \mid \epsilon_t \sim N(z\beta_{1t} + \mathbf{x}_1'\boldsymbol{\beta}_{2t} + \tau\zeta_k + \epsilon_t, \frac{1}{1+\sigma^2}) \tag{2.7}$$

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 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 chapter.

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 (2.5) or the model defined by (2.6) and (2.7) are identifiable.

Remark 2.1: After the reparameterization in (2.4), $\boldsymbol{\beta}_t$, as the ratio of $\boldsymbol{\beta}_t^*$ and $\sqrt{1+\sigma^2}$ is now identifiable. This implies that, in the original formulation of (2.3), 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 1.3 by fitting the model defined by (2.2) and (2.3) without reparameterization, which further confirms the necessity of the reparameterization technique.

2.1.2 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:

$$\text{logit}(P_t) = \log\left(\frac{P_t}{1-P_t}\right) = z\gamma_{1t} + \mathbf{x}_2'\boldsymbol{\gamma}_{2t} + g(\mathbf{R}_{t-1}, \boldsymbol{\gamma}_{3t}) + h(\mathbf{y}_t, \boldsymbol{\gamma}_{4t}), \tag{2.8}$$

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,

$$P(R_0 = r_0, \ldots, R_t = r_t|\mathbf{y}_t, z, \mathbf{x}_2, \boldsymbol{\gamma}) = \prod_{t=0}^{T} P_t^{\mathbf{1}(r_t=1)}(1-P_t)^{\mathbf{1}(r_t=0)}. \tag{2.9}$$

REMARK 2.2: 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 (2.8). The model in (2.9) 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 2.4 for further discussion.)

REMARK 2.3: 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 chapter, 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 2.2.2).

2.2 Bayesian Inference

2.2.1 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

$$\mathcal{L}(\boldsymbol{\theta}|D_c) = \prod_{i=1}^{n}\Big\{f_y(\mathbf{y}_i|z_i, \mathbf{x}_{1i}, k_i, \zeta_{k_i}, \boldsymbol{\epsilon}_i, \mathbf{w}_i, \boldsymbol{\theta})f_{\mathbf{R}|\mathbf{y}}(\mathbf{R}_{iT}|\mathbf{y}_i, z_i, \mathbf{x}_{2i}, \boldsymbol{\theta})\Big\}$$
$$= \prod_{i=1}^{n}\Big\{\prod_{t=0}^{T}\mathbf{1}(w_{it} \geq 0)^{y_{it}}\mathbf{1}(w_{it} < 0)^{1-y_{it}}\frac{1}{\sqrt{2\pi(1-\alpha)}}$$
$$\exp\{-\frac{(w_{it} - z_i\boldsymbol{\beta}_{1t} - \mathbf{x}'_{1i}\boldsymbol{\beta}_{2t} - \tau\zeta_{k_i} - \epsilon_{it})^2}{2(1-\alpha)}\}$$
$$P_{it}{}^{\mathbf{1}(r_{it}=1)}(1-P_{it})^{\mathbf{1}(r_{it}=0)}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\zeta_{k_i}^2}{2}\right)\Big\}\frac{1}{\sqrt{2\pi|\alpha\Sigma|}}\exp\Big\{-\frac{1}{2\alpha}\boldsymbol{\epsilon}'_i\Sigma^{-1}\boldsymbol{\epsilon}_i\Big\}. \quad (2.10)$$

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

$$\mathcal{L}(\boldsymbol{\theta}|D_{\text{obs}}) = \sum_{\mathbf{y}_{\text{mis}}}\int\prod_{i=1}^{n}\Big\{\prod_{t=0}^{T}\mathbf{1}(w_{it} \geq 0)^{y_{it}}\mathbf{1}(w_{it} < 0)^{1-y_{it}}$$
$$\frac{1}{\sqrt{2\pi(1-\alpha)}}\exp\{-\frac{(w_{it} - z_i\boldsymbol{\beta}_{1t} - \mathbf{x}'_{1i}\boldsymbol{\beta}_{2t} - \tau\zeta_{k_i} - \epsilon_{it})^2}{2(1-\alpha)}\}d\mathbf{w}P_{it}{}^{\mathbf{1}(r_{it}=1)}(1-P_{it})^{\mathbf{1}(r_{it}=0)}$$
$$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\zeta_{k_i}^2}{2}\right)d\boldsymbol{\zeta}\Big\}\frac{1}{\sqrt{2\pi|\alpha\Sigma|}}\exp\Big\{-\frac{1}{2\alpha}\boldsymbol{\epsilon}'_i\Sigma^{-1}\boldsymbol{\epsilon}_i\Big\}d\boldsymbol{\epsilon}. \quad (2.11)$$

2.2.2 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

$$\pi(\boldsymbol{\theta}|D_{\text{obs}}) \propto \mathcal{L}(\boldsymbol{\theta}|D_{\text{obs}})\pi(\boldsymbol{\theta}). \quad (2.12)$$

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

Proposition 2.2.1. *Suppose we take $\pi(\boldsymbol{\gamma}) \propto 1$, the joint posterior in (2.12) 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 2.2.1, 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 $1 \leq i \leq n$, $0 \leq t \leq T$. Let $\boldsymbol{\epsilon} = (\epsilon_{\tilde{i}}', i \in I_c)'$, $\mathbf{u}_i = (u_{i0}, \ldots, u_{i,T})'$, $\mathbf{u} = (\mathbf{u}_{\tilde{i}}', i \in I_c)'$, 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 2.2.2. *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 (2.12) 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 (2.11) based on $\tilde{D}_{\text{obs}}$ is given by

$$\begin{aligned}
\ell(\boldsymbol{\theta}|\tilde{D}_{\text{obs}}) = \log \int \prod_{(i,t)\in\tilde{D}_{\text{obs}}} & \mathbf{1}(w_{it} \geq 0)^{y_{it}} \mathbf{1}(w_{it} < 0)^{1-y_{it}} \\
& \frac{1}{\sqrt{2\pi(1-\alpha)}} \exp\{-\frac{(w_{it} - z_i\boldsymbol{\beta}_{1t} - \mathbf{x}'_{1i}\boldsymbol{\beta}_{2t} - \tau\zeta_{k_i} - \epsilon_{it})^2}{2(1-\alpha)}\} d\mathbf{w} \\
& \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\zeta_{k_i}^2}{2}\right) d\boldsymbol{\zeta} \frac{1}{\sqrt{2\pi|\alpha\Sigma|}} \exp\Big\{-\frac{1}{2\alpha}\boldsymbol{\epsilon}'_i\Sigma^{-1}\boldsymbol{\epsilon}_i\Big\} d\boldsymbol{\epsilon} \\
& + \log \prod_{(i,t)\in\tilde{D}_{\text{obs}}} P_{it}{}^{\mathbf{1}(r_{it}=1)}(1-P_{it})^{\mathbf{1}(r_{it}=0)}. \qquad (2.13)
\end{aligned}$$

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 (2.8), 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

$$\pi(\boldsymbol{\gamma}_t) \propto |\mathbf{X}_t^{*\prime}\mathbf{D}_t\mathbf{X}_t^*|^{1/2}, \qquad (2.14)$$

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 (2.14) 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 (2.14), $\mathbf{D}_t$ 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.

2.2.3 Computational Development

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

$$\pi(\boldsymbol{\theta}, \mathbf{y}_{\text{mis}} | D_{\text{obs}}) \propto \mathcal{L}(\boldsymbol{\theta} | D_c) \pi(\boldsymbol{\theta}), \tag{2.15}$$

where $\mathcal{L}(\boldsymbol{\theta}|D_c)$ is defined in (2.10). Thus, the joint posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \alpha, \tau, \rho)$ is written as

$$\begin{aligned}
&\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \alpha, \rho, \tau, \mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, |D_{\text{obs}}) \\
&\propto \prod_{i=1}^{n} \prod_{t=0}^{T} \left\{ \mathbf{1}(w_{it} \geq 0)^{y_{it}} \mathbf{1}(w_{it} < 0)^{1-y_{it}} P_{it}{}^{\mathbf{1}(r_{it}=1)} (1 - P_{it})^{\mathbf{1}(r_{it}=0)} \right\} \\
&\quad (1-\alpha)^{-n(T+1)/2} \prod_{i=1}^{n} \prod_{t=0}^{T} \exp\left\{ -\frac{(w_{it} - z_i \boldsymbol{\beta}_{1t} - \mathbf{x}'_{1i} \boldsymbol{\beta}_{2t} - \tau \zeta_{k_i} - \epsilon_{it})^2}{2(1-\alpha)} \right\} \prod_{i=1}^{n} \prod_{t=0}^{T} \exp\left(-\frac{\zeta_{k_i}^2}{2} \right) \\
&\quad (\alpha)^{-n(T+1)/2} \prod_{i}^{n} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2\alpha} \boldsymbol{\epsilon}'_i \Sigma^{-1} \boldsymbol{\epsilon}_i \right\} \pi(\boldsymbol{\beta}) \pi(\boldsymbol{\gamma}) \pi(\alpha) \pi(\tau) \pi(\rho).
\end{aligned} \tag{2.16}$$

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

$$
\begin{aligned}
&\text{(i)} \quad [\mathbf{y}_{\text{mis}}, \boldsymbol{\gamma} | \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \alpha, \tau, \rho, D_{\text{obs}}]; \quad &&\text{(ii)} \quad [\mathbf{w}, \boldsymbol{\beta} | \mathbf{y}_{\text{mis}}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \alpha, \tau, \rho, D_{\text{obs}}];\\
&\text{(iii)} \quad [\alpha, \rho | \mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \tau, D_{\text{obs}}]; \quad &&\text{(iv)} \quad [\boldsymbol{\epsilon} | \mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \alpha, \tau, \rho, D_{\text{obs}}];\\
&\text{(v)} \quad [\tau | \mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \alpha, \rho, D_{\text{obs}}]; \quad &&\text{(vi)} \quad [\boldsymbol{\zeta} | \mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\epsilon}, \alpha, \tau, \rho, D_{\text{obs}}]. \qquad (2.17)
\end{aligned}
$$

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

$$
\begin{aligned}
&[\mathbf{y}_{\text{mis}}, \boldsymbol{\gamma}, \mathbf{w}, \boldsymbol{\beta} | \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \alpha, \tau, \rho, D_{\text{obs}}] = [\mathbf{y}_{\text{mis}}, \boldsymbol{\gamma} | \boldsymbol{\beta}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \alpha, \tau, \rho, D_{\text{obs}}][\mathbf{w}, \boldsymbol{\beta} | \mathbf{y}_{\text{mis}}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \alpha, \tau, \rho, D_{\text{obs}}]\\
&= [\mathbf{y}_{\text{mis}} | \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \alpha, \tau, \rho, D_{\text{obs}}][\boldsymbol{\gamma} | \mathbf{y}_{\text{mis}}, D_{\text{obs}}][\mathbf{w}, \boldsymbol{\beta} | \mathbf{y}_{\text{mis}}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \alpha, \tau, \rho, D_{\text{obs}}], \qquad (2.18)
\end{aligned}
$$

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:

$$
[\alpha, \rho, \boldsymbol{\epsilon} | \mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \tau, D_{\text{obs}}] = [\alpha, \rho | \mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \tau, D_{\text{obs}}][\boldsymbol{\epsilon} | \mathbf{y}_{\text{mis}}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \alpha, \tau, \rho, D_{\text{obs}}]. \qquad (2.19)
$$

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 (2.5), and $P(r_{ij}|\mathbf{r}_{ij-1}, \mathbf{y}_{ij}, z, \mathbf{x}_2, \boldsymbol{\gamma})$ is given in (2.8). 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 (2.8). Let $\pi(\boldsymbol{\gamma})$ be the Jeffreys prior constructed in Section 2.2.2. 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

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

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 $N\Big(\frac{1}{1-\alpha}\Sigma_1\mu_{1i}, \Sigma_1\Big)$.

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

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

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

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

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., $n_k = \sum_{\{i|k_i=k\}} 1$.

2.2.4 Bayesian Model Assessment

It is of great practical interest to try to 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.*, 2014a).

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 (2.8), $D(\boldsymbol{\psi}) = -2\sum_{i=0}^{n}\sum_{t=0}^{T}[r_{it}\eta_{it}^r - \log(1+\exp(\eta_{it}^r))]$. 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.

$\mathbf{LPML}_{\mathbf{R}|\mathbf{y}}$. 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

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

and the logarithm of the pseudomarginal likelihood is given by

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

Let $\{(\boldsymbol{\psi}_b, \mathbf{y}_{\text{mis},b}, \boldsymbol{\epsilon}_b), b = 1, \ldots, B\}$ denote a Gibbs sample of $(\boldsymbol{\psi}, \mathbf{y}_{\text{mis}}, \boldsymbol{\epsilon})$ from (2.15) and let b represent the b^{th} iteration. A Monte Carlo estimate of CPO_i is given by

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

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

2.3 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: $x_{1i} \sim N(0, 1)$, $x_{2i}|x_{1i} \sim \text{Bernoulli}(1/(1 + \exp(-0.2 - 0.2x_{1i})))$, and the intervention indicator $z_i \sim \text{Bernoulli}(0.5)$. Similar to the HIV prevention behavioral data, we set the total number of visits equal 4. Let $\boldsymbol{\epsilon}^*$ in (2.1) 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

$$\begin{pmatrix} \boldsymbol{\beta}_0^{*\prime} \\ \boldsymbol{\beta}_1^{*\prime} \\ \boldsymbol{\beta}_2^{*\prime} \\ \boldsymbol{\beta}_3^{*\prime} \end{pmatrix} = \begin{pmatrix} -1.0 & 0.5 & 1.0 & 0.4 \\ -1.0 & 0.5 & 1.0 & -0.2 \\ -1.0 & 0.5 & 1.0 & -0.4 \\ -1.0 & 0.5 & 1.0 & -0.6 \end{pmatrix}. \tag{2.20}$$

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

$$\text{logit}(P_{it}) = \gamma_{0t} + x_{1i}\gamma_{1t} + x_{2i}\gamma_{2t} + z_i\gamma_{3t} + \sum_{j=0}^{t-1} R_{ij}\gamma_{4t} + y_{it-1}\gamma_{5t} + y_{it}\gamma_{6t}. \tag{2.21}$$

The missing data mechanism is, therefore, nonignorably missing since P_{it} in (2.21) 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

$$\begin{pmatrix} \boldsymbol{\gamma}_0' \\ \boldsymbol{\gamma}_1' \\ \boldsymbol{\gamma}_2' \\ \boldsymbol{\gamma}_3' \end{pmatrix} = \begin{pmatrix} -2.50 & 0.50 & -0.50 & -0.50 & 0.00 & 0.00 & 0.00 \\ -2.00 & 0.50 & -0.50 & -0.25 & -0.25 & 0.50 & 0.40 \\ -2.80 & 0.50 & -0.50 & 0.25 & -0.60 & 1.30 & 1.70 \\ -2.80 & 0.50 & -0.50 & 0.50 & 0.60 & -0.50 & 1.70 \end{pmatrix}. \tag{2.22}$$

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 (2.22) 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 (2.21) 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 (2.14), 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.

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 $\text{LPML}_{\mathbf{R}|\mathbf{y}}$

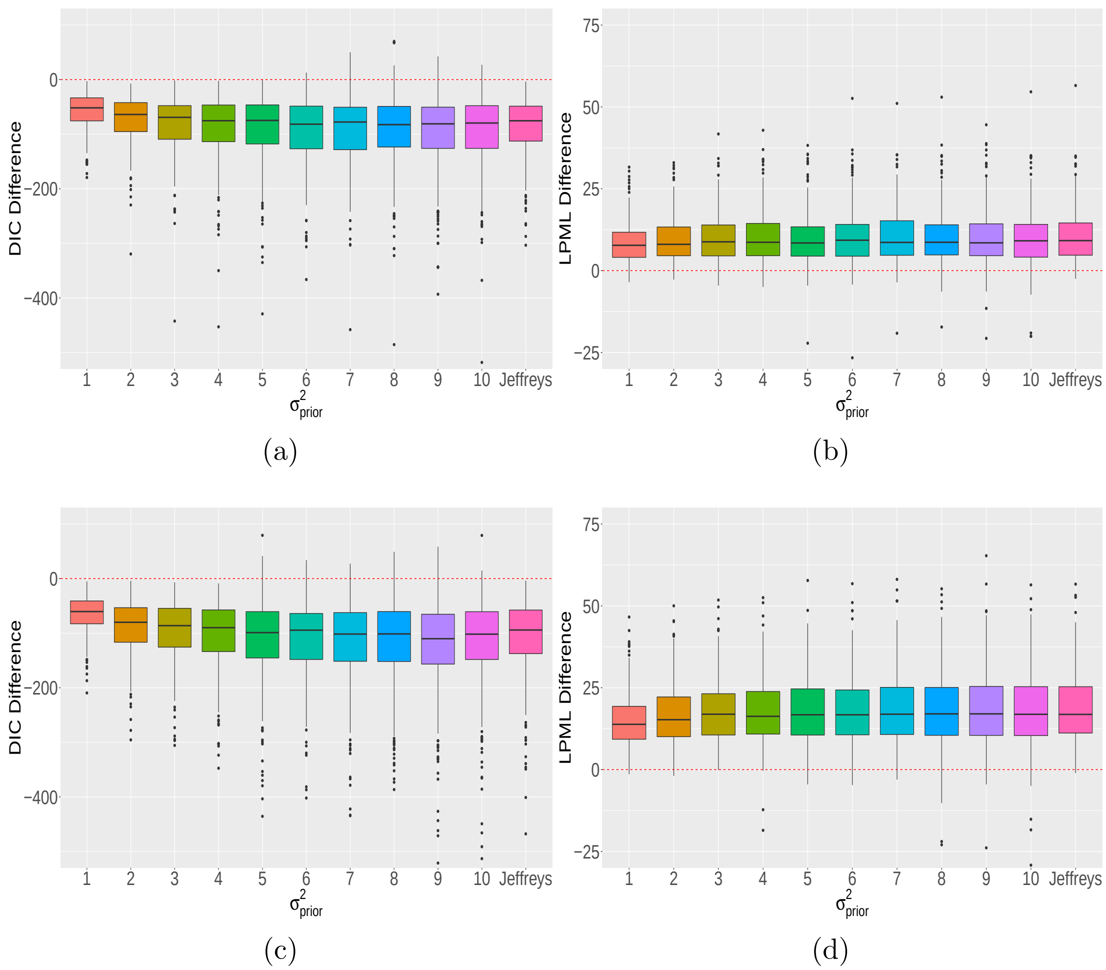


Figure 2.1: Plots of the DIC differences (a) and the LPML differences (b) when the missingness percentages were 5.37%, 10.52%, 11.94%, and 14.18%; and plots of the DIC differences (c) and the LPML differences (d) when the missingness percentages were 5.37%, 10.52%, 11.94%, and 47.14%.

under the nonignorable model minus the $\text{LPML}_{\mathbf{R}|\mathbf{y}}$ under the ignorable model. Figure 2.1 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 2.1, we see that (i) the DIC differences first decrease and then slightly increase as σ^2_{prior} increases (Figure 2.1(a) and Figure 2.1(c)); and (ii) the LPML differences first increase and then slightly decrease as σ^2_{prior} increases (Figure 2.1(b) and Figure 2.1(d)) under both scenarios. Based on Figure 2.1(a) and Figure 2.1(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 2.1(c) and Figure 2.1(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 2.1 and 2.2 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 B.1 and B.2 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.

2.4 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 1.3. 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 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

Table 2.1: 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 2.2: 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

Table 2.3: 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
N(0, 2)	107.06	4755.71	-2397.44
N(0, 3)	114.95	4757.82	-2397.86
N(0, 4)	112.99	4751.86	-2397.70
N(0, 5)	126.66	4748.78	-2397.28
N(0, 6)	132.95	4746.74	-2397.23
N(0, 7)	132.67	4747.22	-2397.23
N(0, 8)	132.94	**4737.61**	**-2396.32**
N(0, 9)	133.47	4745.62	-2397.29
N(0, 10)	140.61	4749.97	-2398.21
Jeffreys Prior	120.18	4750.08	**-2396.64**

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.

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 (2.8). For the nonignorable model, we assumed that $h(\mathbf{y}_t, \boldsymbol{\gamma}_{4t}) = \gamma_{4t1} y_{t-1} + \gamma_{4t2} y_t$ in (2.8) and considered a $N(0, \sigma^2_{prior})$ prior for γ_{4t2} as well as Jeffreys prior for $\boldsymbol{\gamma}_t$ in (2.14). 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 2.3. As exhibited in Table 2.3, 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 2.3 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 2.4-2.6 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 2.4, 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 2.5 were similar to those given in Table 2.4. 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 2.5. 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 2.5 were similar to those given in Table 2.4 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 2.5 in the missing data mechanism, were generally larger than those given in Table 2.4. 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 2.6) were similar to those under the nonignorable model with a $N(0, 8)$ prior (in Table 2.5) for both the binary response model and missing data mechanism, except that the standard deviations for the missing data mechanism in Table 2.6 were slightly smaller. The posterior estimates of ρ, α and τ were similar under the three models.

Table 2.4: 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 2.5: 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 2.6: 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)	—	—	—	—

Chapter 3

Models for Longitudinal Count Response Data

3.1 The Proposed Methods

Suppose there are a total of T visits in a clinical trial. Let y_t denote the measurement for the 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 the control arm and $z = 1$ if the subject is assigned to the intervention arm.

3.1.1 The Model for Longitudinal Counts

As visually displayed in Figure 1.2, the longitudinal count measurements y_t contain lots of zero counts, for $t = 0, \ldots, T$. The conventional Poisson model, therefore, cannot account for the excess zero-count data. Motivated by Figure 1.2, we consider the zero-inflated Poisson model for y_t in this chapter.

Conditioning on the subject-level random effects ϵ_t, we assume that y_t follows the zero-inflated Poisson distribution given by

$$f(y_t|z, \mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \tau, \zeta_k, \epsilon_t) = \begin{cases} \pi_t + (1-\pi_t)e^{-\mu_t} & \text{if } y_t = 0, \\ (1-\pi_t)\frac{\mu_t^{y_t} e^{-\mu_t}}{y_t!} & \text{if } y_t > 0, \end{cases} \tag{3.1}$$

and

$$\begin{aligned} \log(\tfrac{\pi_t}{1-\pi_t}) = & \quad \mathbf{x}_1'\boldsymbol{\beta}_1 + z\gamma_{1t}, \\ \log(\mu_t) = & \quad \mathbf{x}_2'\boldsymbol{\beta}_2 + z\gamma_{2t} + \tau\zeta_k + \epsilon_t, \end{aligned} \tag{3.2}$$

for $t = 0, \ldots, T$, where $\mathbf{x}_1$ and $\mathbf{x}_2$ are two subvectors of $\mathbf{x}$, and $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the vectors of regression coefficients corresponding to $\mathbf{x}_1$ and $\mathbf{x}_2$. 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 effects for all the patients from the k^{th} heath district, $k = 1, \ldots, K$. Let $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \ldots, \zeta_{16})$ be the vector of the sixteen health districts random effects. In (3.2), we 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 the (s, t) entry of Σ being $\rho^{|t-s|}$.

The proposed model is attractive as (i) ϵ_t captures the dependence of the longitudinal measures, y_t, over time; (ii) the time-varying vectors of coefficients γ_{1t} and γ_{2t} allow us to assess the effectiveness of intervention over time; and (iii) the model also takes into account the fact that no obvious variations in count occurs in the control group after the first visit by setting the corresponding z equals 0. We expect γ_{1t} to be greater than 0 given that the number of zero count in the intervention group is greater than the number of zero count in the control group. Furthermore, γ_{2t} should be smaller than 0 considering the decreasing tendency over time of the non-zero counts in the intervention group.

3.1.2 Prequential Multinomial Model for Dropout and Conditional Model for Intermittent Missing

We first propose a prequential multinomial model studying the dropout behavior of each subject. A joint model of all the previous missing status is then developed given that the subject withdraws at certain visit.

Let W denote the visit at which the subject drops out. Then, the possible values of W are $1, 2, \ldots, T, T+1$, where $W = T+1$ indicates that the subject never drops out.

Let

$$p_t = \begin{cases} P(W=1) & \text{if } t=1, \\ P(W=t|W>t-1) & \text{if } 2 \le t \le T. \end{cases} \tag{3.3}$$

Then, we are led to the following theorem.

Theorem 3.1.1. *Assume that p_t is defined in (3.3). Then, we have*

$$P(W=t) = \begin{cases} p_1 & \text{if } t=1, \\ \prod_{\ell=1}^{t-1}(1-p_\ell)p_t & \text{if } 1<t\le T, \\ \prod_{\ell=1}^{T}(1-p_\ell) & \text{if } t=T+1. \end{cases} \tag{3.4}$$

Proof: We use the mathematical induction. It is easy to see that (3.4) holds for $t=1$. For $t=2$, we have $P(W=2) = P(W=2|W>1)P(W>1) = p_2(1-p_1)$. For $2<t<T+1$, we have

$$\begin{aligned} P(W=t) =& P(W=t|W>t-1)P(W>t-1) = p_t P(W>1)\prod_{\ell=2}^{t-1} P(W>\ell|W>\ell-1) \\ =& p_t(1-p_1)\prod_{\ell=2}^{t-1}\{1-P(W=\ell|W>\ell-1)\} = p_t\prod_{\ell=1}^{t-1}(1-p_\ell). \end{aligned}$$

For $t=T+1$, we have

$$P(W=T+1) = P(W>T) = P(W>1)\prod_{\ell=2}^{T} P(W>\ell|W>\ell-1) = \prod_{\ell=1}^{T}(1-p_\ell).$$

This proves the theorem.

Now, we propose logistic regression to model p_t in (3.3):

$$\text{logit}(p_t) = z\alpha_{1t} + \mathbf{x}_3'\boldsymbol{\alpha}_{2t} + h(\boldsymbol{\alpha}_{3t}, \mathbf{y}_t) \tag{3.5}$$

where $\mathbf{x}_3$ is a subvector of $\mathbf{x}$, $\mathbf{y}_t = (y_0, y_1, \ldots, y_t)'$ and $h(\boldsymbol{\alpha}_{3t}, \mathbf{y}_t)$ is a certain linear function of $\mathbf{y}_t$ with $\boldsymbol{\alpha}_{3t}$ being the vector of regression coefficients for $t = 1, 2, \ldots, T$.

Remark 3.1: Instead of using the multinomial logistic regression model for W (Chen *et al.* (2013)) as follows

$$P(W=t) = \begin{cases} \frac{\exp\{z\alpha_{1t}+\mathbf{x}_3'\boldsymbol{\alpha}_{2t}+h(\boldsymbol{\alpha}_{3t},\mathbf{y}_t)\}}{1+\sum_{l=1}^{T}\exp\{z\alpha_{1l}+\mathbf{x}_3'\boldsymbol{\alpha}_{2l}+h(\boldsymbol{\alpha}_{3l},\mathbf{y}_l)\}} & t = 1, \ldots, T, \\ \frac{1}{1+\sum_{l=1}^{T}\exp\{z\alpha_{1l}+\mathbf{x}_3'\boldsymbol{\alpha}_{2l}+h(\boldsymbol{\alpha}_{3l},\mathbf{y}_l)\}} & t = T+1. \end{cases} \tag{3.6}$$

we consider the product of sequential logistic regression in (3.5), which implicitly takes into account the order of longitudinal measurements rather than taking them as a whole in (3.6). Thus, the sequential logistic regression is more appealing in the sense that each p_t depends on the longitudinal values only up to visit t.

Remark 3.2: The two models in (3.5) and (3.6) are exactly the same when $T = 1$. Moreover, if we choose p_t in (3.3) as follows

$$p_t = \frac{\exp\{z\alpha_{1t} + \mathbf{x}_3'\boldsymbol{\alpha}_{2t} + h(\boldsymbol{\alpha}_{3t}, \mathbf{y}_t)\}}{1 + \sum_{l=t}^{T}\exp\{z\alpha_{1l} + \mathbf{x}_3'\boldsymbol{\alpha}_{2l} + h(\boldsymbol{\alpha}_{3l}, \mathbf{y}_l)\}} \quad t = 1, \ldots, T, \tag{3.7}$$

the resulting prequential multinomial model $P(W = t)$ proposed in (3.4) has exactly the same form as the multinomial logistic regression model in (3.6). However, the converse is not true. Thus, the prequential model is more general and favorable.

Remark 3.3: To improve convergence and mixing of the Gibbs sampling algorithm, we introduce the indicator variables $\mathbf{1}(y_t \geq y_t^m)$ in (3.5), where y_t^m is the median of all the nonzero response measurements at visit t, for $t = 1, \ldots, T$. If we assume that $P(W =$

$t|W > t-1)$ depends only on the current and previous values of longitudinal counts, we simply take $h(\boldsymbol{\alpha}_{3t}, \mathbf{y}_t) = \alpha_{3t1}\mathbf{1}(y_{t-1} \geq y_{t-1}^m) + \alpha_{3t2}\mathbf{1}(y_t \geq y_t^m)$. Since y_t is missing, the model in (3.5) is nonignorable. If $h(\boldsymbol{\alpha}_{3t}, \mathbf{y}_t) = \alpha_{3t}\mathbf{1}(y_{t-1} \geq y_{t-1}^m)$, the model in (3.5) is ignorable since y_{t-1} is always observed. However, if $h(\boldsymbol{\alpha}_{3t}, \mathbf{y}_t) = \boldsymbol{\alpha}_{3t}'\mathbf{1}(\mathbf{y}_{t-1} \geq \mathbf{y}_{t-1}^m)$, where $\mathbf{y}_{t-1} = (y_0, y_1, \ldots, y_{t-1})'$, the model in (3.5) might still be nonignorable due to the existence of intermittent missing.

Let R_0 denote the missing status of baseline. The missing indicator R_j at time point j is given by

$$R_j = \begin{cases} 0 & y_j \text{ is observed,} \\ \\ 1 & y_j \text{ is missing,} \quad j = 0, \ldots, T, \end{cases}$$

Assume that the subject drops out at visit t ($W = t$), then we know for sure that the missing indicator $R_j = 1$ with probability one for $j = t, \ldots, T$. Moreover, y_{t-1} has to be observed ($R_{t-1} = 0$) with probability one. Otherwise, W will be smaller than t, which is in conflict with our assumption. Thus, given $W = t > 1$, we only need to model the joint distribution of $(R_0, \ldots, R_{t-2})$. To reduce the number of nuisance parameters, we write the joint distribution of $(R_0, \ldots, R_{t-2})$ through a sequence of one-dimensional conditional distributions. For $t = 1$, $P(R_0 = r_0|W = 1) = 1 - r_0$.

For $t > 1$,

$$\begin{aligned} &P(R_0 = r_0, \ldots, R_{t-2} = r_{t-2}|W = t) = f(r_0, \ldots, r_{t-2}|W = t) \\ =&f(r_0|W = t) \ldots f(r_{t-2}|W = t, r_0, \ldots, r_{t-3}). \end{aligned} \tag{3.8}$$

We assume the conditional distribution of R_j given $W = t$ is given by

$$P(R_j = r_j | W = t, R_0 = r_0, \ldots, R_{j-1} = r_{j-1}) = P(R_j = r_j | R_0 = r_0, \ldots, R_{j-1} = r_{j-1})$$
$$= \begin{cases} q_j^{r_j}(1-q_j)^{1-r_j} & \text{if } 0 \le j \le t-2, \\ 1 - r_j & \text{if } j = t-1, \end{cases} \tag{3.9}$$

where $q_j = P(r_j = 1 | r_0, \ldots, r_{j-1})$ is modeled by a logistic regression:

$$\text{logit}(q_j) = z\phi_{1j} + \mathbf{x}_4' \boldsymbol{\phi}_{2j} + g(\boldsymbol{\phi}_{3j}, \mathbf{R}_{j-1}, \mathbf{y}_j), \tag{3.10}$$

where $\mathbf{R}_{j-1} = (R_0, \ldots, R_{j-1})$, $\mathbf{y}_j = (y_0, y_1, \ldots, y_j)'$ for $j > 1$, and we assume there is no $\mathbf{R}_{j-1}$ for $j = 0$. Let $\mathbf{x}_4$ be a subvector of $\mathbf{x}$, and $g(\boldsymbol{\phi}_{3j}, \mathbf{R}_{j-1}, \mathbf{y}_j)$ is certain linear function of $(\mathbf{R}_{j-1}, \mathbf{y}_j)$ with $\boldsymbol{\phi}_{3j}$ being the vector of regression coefficient for $j = 0, 1, \ldots, t-2$.

We construct the joint distribution of W and R via a marginal and a sequence of one-dimensional conditional distributions,

$$P(W = t, R_0 = r_0, \ldots, R_t = r_t | \mathbf{y}_t, z, \mathbf{x}_1, \mathbf{x}_2, \boldsymbol{\alpha}, \boldsymbol{\phi})$$
$$= \left\{ \prod_{\ell=1}^{t-1} (1 - p_\ell) \right\}^{\mathbf{1}(t>1)} {p_t}^{\mathbf{1}(t<T+1)} \left\{ \prod_{j=0}^{t-2} q_j^{r_j} (1-q_j)^{1-r_j} \right\}^{\mathbf{1}(t>1)} . \tag{3.11}$$

3.1.3 Predictive Probabilities of Dropout and Intermittent Missing

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\phi}, \tau, \sigma^2, \rho)$ denote the collection of all model parameters. We want to compute

$$
\begin{aligned}
& P(\text{Completely Observed}|z, \boldsymbol{\theta}) \\
=& \int P(\text{Completely Observed}|z, \boldsymbol{\theta}, \mathbf{y}, \mathbf{x}) f(\mathbf{y}|z, \boldsymbol{\theta}, \mathbf{x}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}) f(\mathbf{x}|z, \boldsymbol{\theta}) f(\boldsymbol{\zeta}) f(\boldsymbol{\epsilon}|\sigma^2, \rho) d\mathbf{y} d\mathbf{x} d\boldsymbol{\zeta} d\boldsymbol{\epsilon} \\
=& \int P(W = T+1|z, \boldsymbol{\theta}, \mathbf{y}, \mathbf{x}) P(R_0 = 0, \ldots, R_{T-1} = 0|z, \boldsymbol{\theta}, \mathbf{y}, \mathbf{x}, W) \\
& f(\mathbf{y}|z, \boldsymbol{\theta}, \mathbf{x}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}) f(\mathbf{x}|z, \boldsymbol{\theta}) f(\boldsymbol{\zeta}) f(\boldsymbol{\epsilon}|\sigma^2, \rho) d\mathbf{y} d\mathbf{x} d\boldsymbol{\zeta} d\boldsymbol{\epsilon} \\
=& \int \prod_{\ell=1}^{T} (1 - p_\ell) \prod_{j=0}^{T-1} (1 - q_j) f(\mathbf{y}|z, \boldsymbol{\theta}, \mathbf{x}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}) f(\mathbf{x}|z, \boldsymbol{\theta}) f(\boldsymbol{\zeta}) f(\boldsymbol{\epsilon}|\sigma^2, \rho) d\mathbf{y} d\mathbf{x} d\boldsymbol{\zeta} d\boldsymbol{\epsilon},
\end{aligned}
$$

$$
\begin{aligned}
& P(\text{Intermittent Missing Only}|z, \boldsymbol{\theta}) \\
=& P(W = T+1|z, \boldsymbol{\theta}) - P(\text{Completely Observed}|z, \boldsymbol{\theta}) \\
=& \int \prod_{\ell=1}^{T} (1 - p_\ell) \left(1 - \prod_{j=0}^{T-1} (1 - q_j)\right) f(\mathbf{y}|z, \boldsymbol{\theta}, \mathbf{x}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}) f(\mathbf{x}|z, \boldsymbol{\theta}) f(\boldsymbol{\zeta}) f(\boldsymbol{\epsilon}|\sigma^2, \rho) d\mathbf{y} d\mathbf{x} d\boldsymbol{\zeta} d\boldsymbol{\epsilon},
\end{aligned}
$$

$$
\begin{aligned}
& P(\text{Dropout W/ Intermittent Missing }|z, \boldsymbol{\theta}) \\
=& \sum_{t=2}^{T} P(W = t|z, \boldsymbol{\theta})(1 - P(R_0 = 0, \ldots, R_{t-2} = 0|z, \boldsymbol{\theta})) \\
=& \int \sum_{t=2}^{T} \prod_{\ell}^{t-1} (1 - p_\ell) p_t \left(1 - \prod_{j=0}^{t-2} (1 - q_j)\right) f(\mathbf{y}|z, \boldsymbol{\theta}, \mathbf{x}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}) f(\mathbf{x}|z, \boldsymbol{\theta}) f(\boldsymbol{\zeta}) f(\boldsymbol{\epsilon}|\sigma^2, \rho) d\mathbf{y} d\mathbf{x} d\boldsymbol{\zeta} d\boldsymbol{\epsilon},
\end{aligned}
$$

and

$$
\begin{aligned}
& P(\text{Dropout W/O Intermittent Missing}|z, \theta) \\
=& \sum_{t=1}^{T} P(W = t|z, \boldsymbol{\theta}) P(R_0 = 0, \ldots, R_{t-2} = 0|z, \boldsymbol{\theta}) \\
=& \int \left(p_1 + \sum_{t=2}^{T} \prod_{\ell}^{t-1} (1 - p_\ell) p_t \prod_{j=0}^{t-2} (1 - q_j)\right) f(\mathbf{y}|z, \boldsymbol{\theta}, \mathbf{x}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}) f(\mathbf{x}|z, \boldsymbol{\theta}) f(\boldsymbol{\zeta}) f(\boldsymbol{\epsilon}|\sigma^2, \rho) d\mathbf{y} d\mathbf{x} d\boldsymbol{\zeta} d\boldsymbol{\epsilon}
\end{aligned}
$$

3.2 Bayesian Inference

3.2.1 The Likelihood Function

Suppose there are n subjects. Let $\mathbf{y}_{\text{obs}} = (\mathbf{y}'_{0,\text{obs}}, \ldots, \mathbf{y}'_{n,\text{obs}})'$ and $\mathbf{y}_{\text{mis}} = (\mathbf{y}'_{0,\text{mis}}, \ldots, \mathbf{y}'_{n,\text{mis}})'$, where $(\mathbf{y}_{i,\text{obs}}, \mathbf{y}_{i,\text{mis}})$ are the observed and missing measures for the i^{th} subject. Let $\mathbf{y}_i = (y_{i0}, \ldots, y_{iT})'$, $\boldsymbol{\epsilon}_i = (\epsilon_{i0}, \ldots, \epsilon_{iT})'$, $W_i = w_i$ denote the drop-out time for the i^{th} subject, and $\mathbf{R}_i$ denote the collection of all missing indicators $\mathbf{R}_i = (R_{i0}, \ldots, R_{i(w_i-2)})$. Denote by $D_c = \{\mathbf{y}_i, z_i, \mathbf{x}_i, k_i, \zeta_{k_i}, \boldsymbol{\epsilon}_i, W_i, \mathbf{R}_i, i = 1, \ldots, n\}$ the set of the complete data and $D_{\text{obs}} = \{\mathbf{y}_{i,\text{obs}}, z_i, \mathbf{x}_i, k_i, W_i, \mathbf{R}_i, i = 1, \ldots, n\}$ the set of the observed data.

We further assume $(z_i, \mathbf{X}_i)$ are completely observed, for all $i = 1, \ldots, n$. Denote by f_y, $f_{\mathbf{W}}$, $f_{\mathbf{R}|\mathbf{W}}$ the marginal density of $\mathbf{y}$, $\mathbf{W}$, and the conditional density of $\mathbf{R}$ given $\mathbf{W}$. Recall that $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\phi}, \tau, \sigma^2, \rho)$ is the collection of all parameters.

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

$$\mathcal{L}(\boldsymbol{\theta}|D_c) = \prod_{i=1}^{n} f_y(y_{i0}, \ldots, y_{iT}|z_i, \mathbf{x}_i, k_i, \zeta_{k_i}, \boldsymbol{\epsilon}_i, \boldsymbol{\theta}) f_{\mathbf{W}}(w_i|z_i, \mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\theta})$$

$$f_{\mathbf{R}|\mathbf{W}}(r_{i0}, \ldots, r_{iw_i-2}|z_i, \mathbf{x}_i, \mathbf{y}_i, w_i, \boldsymbol{\theta}) = \prod_{i=1}^{n} \left(\prod_{t=0}^{T} \left[\left\{\pi_{it} + (1-\pi_{it})e^{-\mu_{it}}\right\}^{\mathbf{1}(y_{it}=0)} \right.\right.$$
$$\left. \left\{(1-\pi_{it})\frac{\mu_{it}^{y_{it}} e^{-\mu_{it}}}{y_{it}!}\right\}^{\mathbf{1}(y_{it}>0)} \right] \prod_{t=0}^{T} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\zeta_{k_i}^2}{2}\right) \frac{1}{\sqrt{2\pi|\sigma^2\Sigma|}} \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{\epsilon}_i'\Sigma^{-1}\boldsymbol{\epsilon}_i\right\}$$
$$\left. \left\{\prod_{\ell=1}^{w_i-1}(1-p_{i\ell})\right\}^{\mathbf{1}(w_i>1)} {p_{iw_i}}^{\mathbf{1}(w_i<T+1)} \left\{\prod_{j=0}^{w_i-2} q_{ij}^{r_{ij}}(1-q_{ij})^{1-r_{ij}}\right\}^{\mathbf{1}(w_i>1)} \right). \tag{3.12}$$

After integrating out the missing longitudinal measures $\mathbf{y}_{i,\mathrm{mis}}$, ζ_{k_i} and the random vector $\boldsymbol{\epsilon}_i$, the observed data likelihood function is given by

$$\mathcal{L}(\boldsymbol{\theta}|D_{\mathrm{obs}}) = \prod_{i=1}^{n} \int \prod_{t=0}^{T} \left[\{\pi_{it} + (1-\pi_{it})e^{-\mu_{it}}\}^{\mathbf{1}(y_{it}=0)} \left\{ (1-\pi_{it}) \frac{\mu_{it}^{y_{it}} e^{-\mu_{it}}}{y_{it}!} \right\}^{\mathbf{1}(y_{it}>0)} \right]$$

$$\prod_{t=0}^{T} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\zeta_{k_i}^2}{2} \right) \frac{1}{\sqrt{2\pi|\sigma^2\Sigma|}} \exp\left\{ -\frac{1}{2\sigma^2} \boldsymbol{\epsilon}_i' \Sigma^{-1} \boldsymbol{\epsilon}_i \right\} d\zeta_{k_i} d\boldsymbol{\epsilon}_i$$

$$\left\{ \prod_{\ell=1}^{w_i-1} (1-p_{i\ell}) \right\}^{\mathbf{1}(w_i>1)} p_{iw_i}{}^{\mathbf{1}(w_i<T+1)} \left\{ \prod_{j=0}^{w_i-2} q_{ij}^{r_{ij}} (1-q_{ij})^{1-r_{ij}} \right\}^{\mathbf{1}(w_i>1)} d\mathbf{y}_{i,\mathrm{mis}}. \qquad (3.13)$$

3.2.2 Adjusted Intervention Effect

One inferential research goal is to compare the treatment effects in the longitudinal clinical trial. This is not trivial under the zero-inflated Poisson model in (3.2), where the intervention effects are involved separately in both binary and Poisson components. To be more specific, γ_{1t} reflects the partial intervention effects from the excess zero-count measurements while γ_{2t} indicates the partial intervention effects from the Poisson measurements. To assess the overall intervention effects, we need to develop approaches to effectively integrate the two partial intervention effects.

We first note that the intervention effects for simple Poisson model can be obtained by $\log\{E(\mathbf{y}_t|z=1,\mathbf{x})/E(\mathbf{y}_t|z=0,\mathbf{x})\} = \gamma_t$. Inspired by this idea and following Zhang *et al.* (2014b), we provide the following definition of the overall intervention effects under the zero-inflated Poisson,

Definition 3.2.1. *Let γ_t denote the overall intervention effects at visit t adjusting for the covariates and health centers, then we have*

$$\gamma_t \doteq \log \left\{ \frac{\int E(\mathbf{y}_t|z=1,\mathbf{x},\zeta_k,\epsilon_t) f(\zeta_k) f(\epsilon_t|\sigma^2) d\zeta_k d\epsilon_t f(\mathbf{x}|z=1) d\mathbf{x}}{\int E(\mathbf{y}_t|z=0,\mathbf{x},\zeta_k,\epsilon_t) f(\zeta_k) f(\epsilon_t|\sigma^2) d\zeta_k d\epsilon_t f(\mathbf{x}|z=0) d\mathbf{x}} \right\}.$$

After some algebra, we obtain the explicit form of the overall intervention effects.

Proposition 3.2.2.

$$\gamma_t = \log\left\{\frac{\int\left(1-\frac{\exp(\mathbf{x}_1'\boldsymbol{\beta}_1+\gamma_{1t})}{1+\exp(\mathbf{x}_1'\boldsymbol{\beta}_1+\gamma_{1t})}\right)\exp(\mathbf{x}_2'\boldsymbol{\beta}_2+\gamma_{2t})f(\mathbf{x}|z=1)d\mathbf{x}}{\int\left(1-\frac{\exp(\mathbf{x}_1'\boldsymbol{\beta}_1)}{1+\exp(\mathbf{x}_1'\boldsymbol{\beta}_1)}\right)\exp(\mathbf{x}_2'\boldsymbol{\beta}_2)f(\mathbf{x}|z=0)d\mathbf{x}}\right\},$$

where $f(\mathbf{x}|z)$ is the conditional empirical distribution of baseline covariates given intervention indicator.

Proof:

$$\gamma_t = \log\left\{\frac{\int\left(1-\pi_{t|z=1}\right)\lambda_{t|z=1}f(\zeta_k)f(\epsilon_t|\sigma^2)d\zeta_k d\epsilon_t f(\mathbf{x}|z=1)d\mathbf{x}}{\int\left(1-\pi_{t|z=0}\right)\lambda_{t|z=0}f(\zeta_k)f(\epsilon_t|\sigma^2)d\zeta_k d\epsilon_t f(\mathbf{x}|z=0)d\mathbf{x}}\right\}$$

$$= \log\left\{\frac{\int\left(1-\pi_{t|z=1}\right)\exp(\mathbf{x}_2'\boldsymbol{\beta}_2+\gamma_{2t}+\tau\zeta_k+\epsilon_t)\frac{1}{\sqrt{2\pi}}\exp(-\frac{\zeta_k^2}{2})}{\int\left(1-\pi_{t|z=1}\right)\exp(\mathbf{x}_2'\boldsymbol{\beta}_2+\tau\zeta_k+\epsilon_t)\frac{1}{\sqrt{2\pi}}\exp(-\frac{\zeta_k^2}{2})}\right.$$

$$\left.\frac{\frac{1}{\sqrt{2\pi\sigma}}\exp(-\frac{\epsilon_t^2}{2\sigma})d\zeta_k d\epsilon_t f(\mathbf{x}|z=1)d\mathbf{x}}{\frac{1}{\sqrt{2\pi\sigma}}\exp(-\frac{\epsilon_t^2}{2})d\zeta_k d\epsilon_t f(\mathbf{x}|z=0)d\mathbf{x}}\right\}$$

$$= \log\left\{\frac{\int\left(1-\pi_{t|z=1}\right)\exp(\mathbf{x}_2'\boldsymbol{\beta}_2+\gamma_{2t})\exp(-\frac{\tau^2+\sigma^2}{2\sigma})f(\mathbf{x}|z=1)d\mathbf{x}}{\int\left(1-\pi_{t|z=1}\right)\exp(\mathbf{x}_2'\boldsymbol{\beta}_2)\exp(-\frac{\tau^2+\sigma^2}{2})f(\mathbf{x}|z=0)d\mathbf{x}}\right\}$$

$$= \log\left\{\frac{\int\left(1-\frac{\exp(\mathbf{x}_1'\boldsymbol{\beta}_1+\gamma_{1t})}{1+\exp(\mathbf{x}_1'\boldsymbol{\beta}_1+\gamma_{1t})}\right)\exp(\mathbf{x}_2'\boldsymbol{\beta}_2+\gamma_{2t})f(\mathbf{x}|z=1)d\mathbf{x}}{\int\left(1-\frac{\exp(\mathbf{x}_1'\boldsymbol{\beta}_1)}{1+\exp(\mathbf{x}_1'\boldsymbol{\beta}_1)}\right)\exp(\mathbf{x}_2'\boldsymbol{\beta}_2)f(\mathbf{x}|z=0)d\mathbf{x}}\right\}.$$

3.2.3 Prior and Posterior Distributions

We assume that the joint prior density can be expressed by

$$\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\beta})\pi(\boldsymbol{\gamma})\pi(\boldsymbol{\alpha})\pi(\boldsymbol{\phi})\pi(\tau)\pi(\sigma^2)\pi(\rho). \tag{3.14}$$

The joint posterior given the observed data D_{obs} is thus written as follows

$$\pi(\boldsymbol{\theta}|D_{\text{obs}}) \propto \mathcal{L}_o(\boldsymbol{\theta}|D_{\text{obs}})\pi(\boldsymbol{\theta}). \tag{3.15}$$

Since the non-informative priors are considered comparatively more objective and the use of them has become quite a routine in Bayesian analysis, we begin our study of the propriety of the posterior distribution by the non-informative uniform prior.

Result 3.2.3. *Suppose we take* $\pi(\boldsymbol{\beta}_1, \boldsymbol{\gamma}_1, \boldsymbol{\alpha}, \boldsymbol{\phi}) \propto 1$*, the joint posterior in (3.15) is improper regardless of whether* $\pi(\boldsymbol{\beta}_2, \boldsymbol{\gamma}_2, \tau, \sigma^2, \rho)$ *is proper or improper.*

If we specify uniform priors for the Poisson regression coefficients, i.e., $\pi(\boldsymbol{\beta}_2, \boldsymbol{\gamma}_2, \tau, \sigma^2, \rho) \propto 1$, the joint posterior in (3.15) is proper, conditioning on the other parameters. However, according to Theorem 2.1 established by Chen and Shao (2001), we can show that the joint posterior is improper if $\pi(\boldsymbol{\beta}_1, \boldsymbol{\gamma}_1) \propto 1$. In particular, if the number of excess zeros is not large enough and the Poisson mean is small, it is ambiguous whether the zeros come from the Poisson or binary model. Therefore, proper priors should be chosen for $(\boldsymbol{\beta}_1, \boldsymbol{\gamma}_1)$. Other causes of the impropriety of the posterior distribution are $(\boldsymbol{\alpha}, \boldsymbol{\phi})$, which involve the coefficients of the missing response measures $\mathbf{y}$. According to Theorem 1 in Huang *et al.* (2005), the joint posterior distribution is improper if the linear predictors in (3.5) and (3.10) include intercepts and $\pi(\boldsymbol{\alpha}, \boldsymbol{\phi}) \propto 1$. We are led to the following result.

Result 3.2.4. *Suppose we take* $\pi(\boldsymbol{\beta}_1)$*,* $\pi(\boldsymbol{\gamma}_1)$*,* $\pi(\boldsymbol{\alpha})$*, and* $\pi(\boldsymbol{\phi})$ *to be proper priors, and specify uniform priors for the other variables. Under some mild conditions of* $(z, \mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_4, \mathbf{y})$*, the joint posterior in (3.15) is proper.*

We consider the modified Jeffreys prior for $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ such that the logarithm of the joint likelihood function of the parameters does not involve any missing data, and is thus analytically attractive:

$$\ell(\boldsymbol{\theta}|\tilde{\mathbf{D}}_{\text{obs}}) = \sum_{\mathbf{i}\in\tilde{\mathbf{D}}_{\text{obs}}} \log \int \mathbf{f_y}(\mathbf{y_{i0}}, \ldots, \mathbf{y_{iT}}|\mathbf{z_i}, \mathbf{x_i}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\epsilon})\mathbf{f}(\zeta_{\mathbf{k_i}}|\tau)\mathbf{d}\zeta_{\mathbf{k_i}}\mathbf{f}(\boldsymbol{\epsilon}_\mathbf{i}|\sigma^\mathbf{2}, \rho)\mathbf{d}\boldsymbol{\epsilon}_\mathbf{i} +$$

$$\sum_{i\in\tilde{D}_{\text{obs}}} \log f_{\mathbf{W}}(w_i|z_i, \mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\alpha}) + \sum_{i\in\tilde{D}_{\text{obs}}} \log f_{\mathbf{R}|\mathbf{W}}(r_{i0}, \ldots, r_{iw_i-2}|z_i, \mathbf{x}_i, \mathbf{y}_i, w_i, \boldsymbol{\phi}), \qquad (3.16)$$

where $\tilde{D}_{\text{obs}}$ is certain observed subset. For $\boldsymbol{\alpha}_t$ at visit t, we use a different observed subset to construct the prior, aiming to utilize as many observations as possible. If we take

$h(\boldsymbol{\alpha}_{3t}, \mathbf{y}_t) = \alpha_{3t1}\mathbf{1}(y_{t-1} \geq y_{t-1}^m) + \alpha_{3t2}\mathbf{1}(y_t \geq y_t^m)$ in (3.5), the log-likelihood of $\boldsymbol{\alpha}_t$ based on the subset data is given by

$$\ell(\boldsymbol{\alpha}_\mathbf{t}|\mathbf{D_c}) = \sum_{\mathbf{i=1}}^{\mathbf{n}} \log\{\mathbf{f_W}(\mathbf{w_i}|\mathbf{z_i}, \mathbf{x_i}, \mathbf{y_i}, \boldsymbol{\alpha})^{\mathbf{1(r_{it-1}=0)1(r_{it}=0)}}\}$$
$$= \sum_{i=1}^{n} \mathbf{1}(r_{it-1} = 0)\mathbf{1}(r_{it} = 0) \left\{ \mathbf{1}(w_i > 1) \sum_{\ell=1}^{w_i-1} \log(1 - p_{i\ell}) + \mathbf{1}(w_i < T + 1) \log p_{iw_i} \right\}.$$

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

$$\pi(\boldsymbol{\alpha}_t) \propto |\mathbf{X}_t^{*\prime}\mathbf{D}_t\mathbf{X}_t^*|^{1/2}, \tag{3.17}$$

where

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

$|.|$ represents the determinant of a matrix, $\mathbf{X}_{it}^* = (z, \mathbf{x}_3', \mathbf{1}(y_{it} \geq y_t^m))'$ if $t = 0$, and $\mathbf{X}_{it}^* = (z, \mathbf{x}_3', \mathbf{1}(y_{it-1} \geq y_{t-1}^m), \mathbf{1}(y_{it} \geq y_t^m))'$ for $t \geq 1$. Also, in (3.17), $\mathbf{D}_t$ 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 α_t is proper. The Jeffreys priors for $\boldsymbol{\phi}_j$ in (3.10) can be derived in the same way.

To summarize, we assume uniform priors for $(\boldsymbol{\beta}_2, \gamma_2, \rho)$, the Jeffreys priors for $(\boldsymbol{\alpha}_t, \boldsymbol{\phi}_j)$, g-priors (Zellner, 1986) for $(\boldsymbol{\beta}_1, \boldsymbol{\gamma}_1)$, the Inverse Gamma distribution for σ^2, and a truncated normal prior for τ.

3.2.4 Computational Development

Due to the complexity of the likelihood function of the conditional model, it is impossible to derive the analytical form of the posterior distribution. Thus, we adopt the Markov chain Monte Carlo (MCMC) methods to sample from the posterior distribution in (3.15).

To simplify the distribution function in (3.1), we introduce the latent random variable b_{it}, with $b_{it} = 1$ indicating the zero count comes from binary distribution and $b_{it} = 0$ if the underlying distribution of the zero count is Poisson. Moreover, we let $b_{it} = 0$ if $y_{it} > 0$. Let $\mathcal{O} = \{y_{it} | y_{it} = 0, t = 0, \ldots, T; i = 1, \ldots, n\}$ denote the set of all zero counts. The conditional distribution of b_{it} given $y_{it} = 0$ is given by

$$b_{it} | y_{it} = 0 \sim Bernoulli\left(\frac{\pi_{it}}{\pi_{it} + (1-\pi_{it})e^{-\mu_{it}}}\right).$$

We then write the likelihood function for $\boldsymbol{\theta}$ as follows

$$\mathcal{L}(\boldsymbol{\theta}|D_c) \propto \prod_{y_{it}\in\mathcal{O}} \pi_{it}^{b_{it}}[(1-\pi_{it})e^{-\mu_{it}}]^{1-b_{it}} \prod_{y_{it}\notin\mathcal{O}} (1-\pi_{it})e^{-\mu_{it}}\mu_{it}^{y_{it}}/y_{it}! \prod_{\{(i,t):b_{it}=0\}} f(\zeta_{k_i}|\tau)$$
$$f(\boldsymbol{\epsilon}_i|\sigma^2,\rho) \prod_{i=1}^{n} f_{\mathbf{W}}(w_i|z_i, \mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\alpha}) f_{\mathbf{R}|\mathbf{W}}(r_{i0}, \ldots, r_{it-2}|z_i, \mathbf{x}_i, \mathbf{y}_i, w_i, \boldsymbol{\phi}). \quad (3.18)$$

The Gibbs sampling algorithm is applied in the order of: (i) $[\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau, \boldsymbol{\zeta}, \sigma^2, \rho, \boldsymbol{\epsilon}, \mathbf{y}_{\text{mis}}|D_{\text{obs}}]$; (ii) $[\boldsymbol{\alpha}|\mathbf{y}_{\text{mis}}, D_{\text{obs}}]$; and (iii) $[\boldsymbol{\phi}|\mathbf{y}_{\text{mis}}, D_{\text{obs}}]$.

For (i), we run a sub-Gibbs sampling algorithm to sample $[\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau, \boldsymbol{\zeta}, \sigma^2, \rho, \boldsymbol{\epsilon}, \mathbf{y}_{\text{mis}}|D_{\text{obs}}]$ from the joint conditional distribution. The sub-Gibbs sampling algorithm requires to sample from the following conditional posterior distributions in turns:

(ia) $[\mathbf{y}_{\text{mis}}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau, \boldsymbol{\zeta}, \sigma^2, \rho, \boldsymbol{\epsilon}, D_{\text{obs}}]$;

(ib) $[\boldsymbol{\beta}_2, \boldsymbol{\gamma}_2, \tau, \boldsymbol{\zeta}, \sigma^2, \rho, \boldsymbol{\epsilon}|\boldsymbol{\beta}_1, \gamma_1, \mathbf{y}_{\text{mis}}, D_{\text{obs}}]$; and

(ic) $[\boldsymbol{\beta}_1, \gamma_1|\boldsymbol{\beta}_2, \boldsymbol{\gamma}_2, \tau, \boldsymbol{\zeta}, \sigma^2, \rho, \boldsymbol{\epsilon}, \mathbf{y}_{\text{mis}}, D_{\text{obs}}]$.

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

Step (ia). The most challenging part of the sub-Gibbs sampling is $\mathbf{y}_{\text{mis}}$ in (ia). Due to the involvement of $\mathbf{y}$ in the likelihood functions of $\mathbf{W}$ and $\mathbf{R}$, the posterior distribution of $\mathbf{y}$ is no longer zero-inflated Poisson. To generate $\mathbf{y}_{\text{mis}}$, we use the following approach.

(I). For each missing response $y_{it,\text{mis}}$, we introduce the kernel function $q_{it,k}$ as follows

$$q_{it,k} = \{\pi_{it} + (1-\pi_{it})e^{-\mu_{it}}\}^{\mathbf{1}(k=0)}\{(1-\pi_{it})e^{-\mu_{it}}\mu_{it}^k/k!\}^{\mathbf{1}(k>0)}$$
$$P(W_i = w_i | y_{it} = k) P(R_{i0} = r_{i0}, \cdots, R_{i(w_i-2)} = r_{i(w_i-2)} | y_{it} = k, W_i = w_i),$$

where it refers to the t^{th} visit for the i^{th} observation, and $k = 0, 1, \ldots$.

(II). We then find the stopping time S_t such that $\frac{q_{it,S_t+1}}{\sum_{l=0}^{S_t} q_{it,l}} < \epsilon$, where ϵ is a small enough value. Let $p_{it,k} = \frac{q_{it,k}}{\sum_{l=0}^{S_t} q_{it,l}}$, for $k = 0, 1, \ldots, S_t$.

(III). Generate $u \sim U(0,1)$

for all $0 \le k \le S_t$ **do**

  if $u < p_{it,k}$ **then**

    $y_{it,\text{mis}} = k$;

    Exit;

  else

    $u = u - p_{it,k}$;

  end if

end for

Remark 3.4: Since we set $h(\boldsymbol{\alpha}_{3t}, \mathbf{y}_t) = \alpha_{3t1}\mathbf{1}(y_{t-1} \ge y_{t-1}^m) + \alpha_{3t2}\mathbf{1}(y_t \ge y_t^m)$ in (3.5), and $g(\boldsymbol{\phi}_{3j}, \mathbf{R}_{j-1}, \mathbf{y}_j) = \boldsymbol{\phi}_{3j1}'\mathbf{R}_{j-1} + \phi_{3j2}\mathbf{1}(y_{j-1} \ge y_{j-1}^m) + \phi_{3j3}\mathbf{1}(y_j \ge y_j^m)$ in (3.10), we can actually derive the exact probability of $p_{it,k}$ in (II):

$$p_{it,k} = q_{it,k}/(a_{it}c_{it} + (1-a_{it})d_{it}),$$

where $a_{it} = \pi_{it} + (1-\pi_{it})e^{-\mu_{it}}(1+\mu_{it}+\cdots+\frac{\mu_{it}{}^{y_t^m-1}}{(y_t^m-1)!})$, $c_{it} = P(W_i = w_i | y_{it} < y_t^m)P(R_{i0} = r_{i0}, \cdots, R_{i(w_i-2)} = r_{i(w_i-2)} | y_{it} < y_t^m, W_i = w_i)$, and $d_{it} = P(W_i = w_i | y_{it} \geq y_t^m)P(R_{i0} = r_{i0}, \cdots, R_{i(w_i-2)} = r_{i(w_i-2)} | y_{it} \geq y_t^m, W_i = w_i)$.

Step (ib). For sampling $(\boldsymbol{\beta}_2, \boldsymbol{\gamma}_2, \tau, \boldsymbol{\zeta}, \sigma^2, \rho, \boldsymbol{\epsilon})$, we use the hierarchical centering technique provided in Chen *et al.* (2012), which greatly improves the convergence of the Gibbs sampler. A hierarchically centered reparameterization is given by

$$\boldsymbol{\eta}_i = \mathbf{X}_{2i}\boldsymbol{\beta}_2 + \mathbf{Z}_i\boldsymbol{\gamma}_2 + \tau\zeta_{k_i}\mathbf{1} + \boldsymbol{\epsilon}_i, \tag{3.19}$$

where $\boldsymbol{\gamma}_2 = (\gamma_{20}, \ldots, \gamma_{2T})'$ is a vector of length $(T+1)$, $\mathbf{Z}_i$ is a $(T+1)\times(T+1)$ diagonal matrix with diagonal element being z_i, $\boldsymbol{\beta}_2$ is a vector of length p_2, $\mathbf{X}_{2i}$ is a $(T+1)\times p_2$ matrix with all the row vectors equal $\mathbf{x}_{2i}'$, $\mathbf{1}$ is a $(T+1)$ length vector of ones, and $\boldsymbol{\eta}_i = (\eta_{i0}, \ldots, \eta_{iT})'$ is also a vector of length $(T+1)$.

The reparameterized posterior for $(\boldsymbol{\beta}_2, \boldsymbol{\gamma}_2, \tau, \boldsymbol{\zeta}, \sigma^2, \rho, \boldsymbol{\eta})$ is written as

$$\pi(\boldsymbol{\beta}_2, \boldsymbol{\gamma}_2, \tau, \boldsymbol{\zeta}, \sigma^2, \rho, \boldsymbol{\eta} | y_{\text{mis}}, D_{\text{obs}}) \propto \prod_{i=1}^{n}\prod_{t=0}^{T}\Big[\{\pi_{it} + (1-\pi_{it})\exp(-\exp(\eta_{it}))\}^{\mathbf{1}(y_{it}=0)}$$
$$\{\exp\{y_{it}\eta_{it} - \exp(\eta_{it}) - \log(y_{it}!)\}^{\mathbf{1}(y_{it}>0)}\Big]\prod_{i=1}^{n}\prod_{t=0}^{T}\exp\left(-\frac{\zeta_{k_i}^2}{2}\right)\sigma^{2-\frac{n(T+1)}{2}}|\Sigma|^{-n/2}$$
$$\prod_{i}^{n}\exp\Big\{-\frac{1}{2\sigma^2}(\boldsymbol{\eta}_i - \mathbf{X}_{2i}\boldsymbol{\beta}_2 - \mathbf{Z}_i\boldsymbol{\gamma}_2 - \tau\zeta_{k_i}\mathbf{1})'\Sigma^{-1}(\boldsymbol{\eta}_i - \mathbf{X}_{2i}\boldsymbol{\beta}_2 - \mathbf{Z}_i\boldsymbol{\gamma}_2 - \tau\zeta_{k_i}\mathbf{1})\Big\}$$
$$\prod_{i=1}^{n}\left[\left\{\prod_{\ell=1}^{w_i-1}(1-p_{i\ell})\right\}^{\mathbf{1}(w_i>1)} p_{iw_i}{}^{\mathbf{1}(w_i<T+1)}\left\{\prod_{j=0}^{w_i-2}q_{ij}^{r_{ij}}(1-q_{ij})^{1-r_{ij}}\right\}^{\mathbf{1}(w_i>1)}\right]$$
$$\pi(\tau)\pi(\sigma^2)\pi(\rho)\pi(\boldsymbol{\beta}_2)\pi(\boldsymbol{\gamma}_2) \tag{3.20}$$

The Gibbs sampler for sampling from the reparameterized posterior $\pi(\boldsymbol{\beta}_2, \boldsymbol{\gamma}_2, \tau, \boldsymbol{\zeta}, \sigma^2, \rho, \boldsymbol{\eta})$ requires the following steps:

The Hierarchical Centering Approach

(I). Draw $\boldsymbol{\eta} = (\boldsymbol{\eta}_1', \ldots, \boldsymbol{\eta}_{n_0}')'$ from its conditional posterior distribution using the localized Metropolis Algorithm for $y_{it} = 0$, and adaptive rejection metropolis algorithm (ARMS)(Gilks and Wild (1992)) for $y_{it} = 1$,

$$\pi(\boldsymbol{\eta}|\boldsymbol{\beta}_2, \boldsymbol{\gamma}_2, \tau, \boldsymbol{\zeta}, \sigma^2, \rho, \mathbf{y}_{\text{mis}}, D_{\text{obs}}) \propto$$
$$\prod_{i=1}^{n}\prod_{t=0}^{T}\Big[\{\pi_{it} + (1-\pi_{it})\exp(-\exp(\eta_{it}))\}^{\mathbf{1}(y_{it}=0)}$$
$$\{\exp\{y_{it}\eta_{it} - \exp(\eta_{it}) - \log(y_{it}!)\}^{\mathbf{1}(y_{it}>0)}\Big]\prod_{i}^{n}\exp\Big\{-\frac{1}{2\sigma^2}(\boldsymbol{\eta}_i - \mathbf{X}_{2i}\boldsymbol{\beta}_2 - \mathbf{Z}_i\boldsymbol{\gamma}_2$$
$$-\tau\zeta_{k_i}\mathbf{1})'\Sigma^{-1}(\boldsymbol{\eta}_i - \mathbf{X}_{2i}\boldsymbol{\beta}_2 - \mathbf{Z}_i\boldsymbol{\gamma}_2 - \tau\zeta_{k_i}\mathbf{1})\Big\}$$

(II). Assume uniform priors for $\tilde{\boldsymbol{\beta}}_2 = (\boldsymbol{\beta}_2', \boldsymbol{\gamma}_2')'$. Let $\tilde{\mathbf{X}}_{2i} = (\mathbf{X}_{2i}, \mathbf{Z}_i)$ be a $(T+1)\times(p_2+T+1)$ vector. We then draw $\tilde{\boldsymbol{\beta}}_2$ from a normal distribution

$$N\left(\left(\sum_{i=1}^{n}\tilde{\mathbf{X}}_{2i}'\Sigma^{-1}\tilde{\mathbf{X}}_{2i}\right)^{-1}\sum_{i=1}^{n}\tilde{\mathbf{X}}_{2i}'\Sigma^{-1}\left(\boldsymbol{\eta}_i - \tau\zeta_{k_i}\mathbf{1}\right), \sigma^2\left(\sum_{i=1}^{n}\tilde{\mathbf{X}}_{2i}'\Sigma^{-1}\tilde{\mathbf{X}}_{2i}\right)^{-1}\right).$$

(III). Draw σ^2 from its conditional posterior

$$\sigma^2|\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau, \boldsymbol{\zeta}, \rho, \boldsymbol{\eta}, \mathbf{y}_{\text{mis}}, D_{\text{obs}} \sim \mathcal{IG}(\delta^*, \lambda^*),$$

where $\mathcal{IG}(\cdot,\cdot)$ is an Inverse Gamma distribution. We assume an $\mathcal{IG}(\delta_0, \lambda_0)$ prior for σ^2. Therefore, we have $\delta^* = \delta_0 + \frac{n(T+1)}{2}$, and $\lambda^* = \lambda_0 + \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{\eta}_i - \mathbf{X}_{2i}\boldsymbol{\beta}_2 - \mathbf{Z}_i\boldsymbol{\gamma}_2 - \tau\zeta_{k_i}\mathbf{1})'\Sigma^{-1}(\boldsymbol{\eta}_i - \mathbf{X}_{2i}\boldsymbol{\beta}_2 - \mathbf{Z}_i\boldsymbol{\gamma}_2 - \tau\zeta_{k_i}\mathbf{1})$.

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

$$\pi(\tau|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \sigma^2, \rho, \boldsymbol{\eta}, \mathbf{y}_{\text{mis}}, D_{\text{obs}}) \propto \prod_{i}^{n}\Big[\exp\Big\{-\frac{1}{2\sigma^2}(\boldsymbol{\eta}_i - \mathbf{X}_{2i}\boldsymbol{\beta}_2 - \mathbf{Z}_i\boldsymbol{\gamma}_2$$
$$-\tau\zeta_{k_i}\mathbf{1})'\Sigma^{-1}(\boldsymbol{\eta}_i - \mathbf{X}_{2i}\boldsymbol{\beta}_2 - \mathbf{Z}_i\boldsymbol{\gamma}_2 - \tau\zeta_{k_i}\mathbf{1})\Big\}\Big]\pi(\tau).$$

Assume τ follows the truncated normal prior $\tau \sim N(0, 10)\mathbf{1}(\tau > 0)$. Let

$$\mu_{\text{new}} = \frac{\sum_{i=1}^{n} \zeta_{k_i} \mathbf{1}'\Sigma^{-1}(\boldsymbol{\eta}_i - \mathbf{X}_{2i}\boldsymbol{\beta}_2 - \mathbf{Z}_i\boldsymbol{\gamma}_2)}{\sum_{i=1}^{n} \zeta_{k_i}^2 \mathbf{1}'\Sigma^{-1}\mathbf{1}},$$

and

$$\sigma_{\text{new}}^2 = \frac{\sigma^2}{\sum_{i=1}^{n} \zeta_{k_i}^2 \mathbf{1}'\Sigma^{-1}\mathbf{1}}.$$

We then draw τ from the posterior distribution

$$N\left(\frac{\frac{\mu_{\text{new}}}{\sigma_{\text{new}}^2}}{\frac{1}{\sigma_{\text{new}}^2} + \frac{1}{10}}, \frac{1}{\frac{1}{\sigma_{\text{new}}^2} + \frac{1}{10}}\right)\mathbf{1}(\tau > 0).$$

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

$$\pi(\zeta_k|\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau, \sigma^2, \rho, \boldsymbol{\eta}, \mathbf{y}_{\text{mis}}, D_{\text{obs}}) \propto \prod_{\{i:k_i=k\}} \Big[\exp\Big\{ - \frac{1}{2\sigma^2}(\boldsymbol{\eta}_i - \mathbf{X}_{2i}\boldsymbol{\beta}_2 - \mathbf{Z}_i\boldsymbol{\gamma}_2$$

$$- \tau\zeta_{k_i}\mathbf{1})'\Sigma^{-1}(\boldsymbol{\eta}_i - \mathbf{X}_{2i}\boldsymbol{\beta}_2 - \mathbf{Z}_i\boldsymbol{\gamma}_2 - \tau\zeta_{k_i}\mathbf{1})\Big\}\Big] \prod_{t=0}^{T} \prod_{\{i:k_i=k\}} \exp\left(-\frac{\zeta_{k_i}^2}{2}\right).$$

Let

$$\mu_{\text{new}} = \frac{\sum_{\{i:k_i=k\}} \tau\mathbf{1}'\Sigma^{-1}(\boldsymbol{\eta}_i - \mathbf{X}_{2i}\boldsymbol{\beta}_2 - \mathbf{Z}_i\boldsymbol{\gamma}_2)}{\sum_{\{i:k_i=k\}} \tau^2\mathbf{1}'\Sigma^{-1}\mathbf{1}},$$

and

$$\sigma_{\text{new}}^2 = \frac{\sigma^2}{\sum_{\{i:k_i=k\}} \tau^2\mathbf{1}'\Sigma^{-1}\mathbf{1}}.$$

We then draw ζ_k from a $N\left(\frac{\frac{\mu_{\text{new}}}{\sigma_{\text{new}}^2}}{\frac{1}{\sigma_{\text{new}}^2}+n_k}, \frac{1}{\frac{1}{\sigma_{\text{new}}^2}+n_k}\right)$ distribution for $k = 1, \ldots, 16$, where $n_k = (T+1)\sum_{\{i:k_i=k\}} 1$.

(VI). Assume $\rho \sim U(-1, 1)$, the conditional posterior distribution of ρ is given by

$$\pi(\rho|\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau, \boldsymbol{\zeta}, \sigma^2, \boldsymbol{\eta}, \mathbf{y}_{\text{mis}}, D_{\text{obs}}) \propto |\Sigma|^{-n/2} \prod_{i}^{n} \exp\Big\{ - \frac{1}{2\sigma^2}(\boldsymbol{\eta}_i - \mathbf{X}_{2i}\boldsymbol{\beta}_2 -$$

$$\mathbf{Z}_i\boldsymbol{\gamma}_2 - \tau\zeta_{k_i}\mathbf{1})'\Sigma^{-1}(\boldsymbol{\eta}_i - \mathbf{X}_{2i}\boldsymbol{\beta}_2 - \mathbf{Z}_i\boldsymbol{\gamma}_2 - \tau\zeta_{k_i}\mathbf{1})\Big\}$$

Since $-1 < \rho < 1$, we introduce a "de-constraint" transformation to sample ρ (Chen *et al.* (2012)),

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

Thus

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

Since $\pi(\xi|\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau, \boldsymbol{\zeta}, \sigma^2, \boldsymbol{\eta}, \mathbf{y}_{\text{mis}}, D_{\text{obs}})$ is not log-concave, we use the localized Metropolis Algorithm to sample ξ, and then convert back to ρ.

Step (ic). We specify g-priors for $\tilde{\boldsymbol{\beta}}_1 = (\boldsymbol{\beta}_1', \boldsymbol{\gamma}_1')'$,

$$\tilde{\boldsymbol{\beta}}_1 \sim N\left(\mathbf{0}, g\left(\sum \tilde{\mathbf{x}}_{1i}\tilde{\mathbf{x}}_{1i}'\right)^{-1}\right) \quad , \text{ and } \quad \pi(g) = \frac{1}{(1+g)^2},$$

where $\tilde{\mathbf{x}}_{1i} = (\mathbf{x}_{1i}', \mathbf{z}_{it}')'$ be a $(p_1 + T + 1)$ length of vector, where $\mathbf{z}_{it}$ is the $(t+1)^{th}$ column vector of $\mathbf{Z}_i$, where $t = 0, \ldots, T$. Since $g > 0$, we again introduce a "de-constraint" transformation $g = e^{\nu}$ to sample g,

$$\pi(\nu|\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau, \boldsymbol{\zeta}, \sigma^2, \rho, \boldsymbol{\eta}, \mathbf{y}_{\text{mis}}, D_{\text{obs}}) \propto \exp\left\{-\left(\frac{\nu(p_1 + T + 1)}{2} + \frac{\tilde{\boldsymbol{\beta}}_1' \left(\sum \tilde{\mathbf{x}}_{1i}\tilde{\mathbf{x}}_{1i}'\right) \tilde{\boldsymbol{\beta}}_1}{2e^{\nu}}\right)\right\} \cdot (1 + e^{\nu})^{-2} \cdot e^{\nu},$$

and the conditional posterior distribution of $\tilde{\boldsymbol{\beta}}_1$ is given by

$$\pi(\tilde{\boldsymbol{\beta}}_1|\boldsymbol{\beta}_2, \boldsymbol{\gamma}_2, \tau, \boldsymbol{\zeta}, \sigma^2, \rho, \boldsymbol{\eta}, \mathbf{y}_{\text{mis}}, D_{\text{obs}}) \propto \prod_{y_{it} \in \mathcal{O}} \pi_{it}^{b_{it}} (1 - \pi_{it})^{1 - b_{it}} \prod_{y_{it} \notin \mathcal{O}} (1 - \pi_{it}) \exp\left(-\frac{\tilde{\boldsymbol{\beta}}_1' \left(\sum \tilde{\mathbf{x}}_{1i}\tilde{\mathbf{x}}_{1i}'\right) \tilde{\boldsymbol{\beta}}_1}{2g}\right),$$

where π_{it} is established in (3.2). Localized Metropolis Algorithm is used for sampling g and $\tilde{\boldsymbol{\beta}}_1$.

Step (ii). The posterior conditional distribution of $\boldsymbol{\alpha}$ is given by

$$\pi(\boldsymbol{\alpha}|\mathbf{y}_{\text{mis}}, D_{\text{obs}}) \propto \prod_{i=1}^{n} \left\{ \prod_{\ell=1}^{t-1} (1-p_{i\ell}) \right\}^{\mathbf{1}(t\neq T)} p_{it}^{\mathbf{1}(t\neq T+1)} \pi(\boldsymbol{\alpha}),$$

where p_{it} is given in (3.3).

Step (iii). The posterior conditional distribution of $\boldsymbol{\phi}$ is given by

$$\pi(\boldsymbol{\phi}|\mathbf{y}_{\text{mis}}, D_{\text{obs}}) \propto \prod_{i=1}^{n} \left\{ \prod_{j=0}^{w_i-2} q_{ij}^{r_{ij}} (1-q_{ij})^{1-r_{ij}} \right\}^{\mathbf{1}(w_i>1)} \pi(\boldsymbol{\phi})$$

Let $\pi(\boldsymbol{\alpha})$ and $\pi(\boldsymbol{\phi})$ be the Jeffreys priors established in Section 3.2.3. We cannot use adaptive rejection sampling since Jeffreys prior is not log-concave. Therefore, we again use the Localized Metropolis Algorithm.

3.2.5 Bayesian Model Assessment

Similar to Section 2.2.4, we consider two Bayesian model assessment criteria, namely, the DIC relating to the missing data model ($\text{DIC}_{W,\mathbf{R}|\mathbf{y}}$)(Yao *et al.*, 2015; Mason *et al.*, 2012), and the LMPL relating to the missing data model ($\text{LPML}_{W,\mathbf{R}|\mathbf{y}}$) (Zhang *et al.*, 2014a).

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

$\text{DIC}_{W,\mathbf{R}|\mathbf{y}}$. Let $\boldsymbol{\psi} = (\boldsymbol{\alpha}, \boldsymbol{\phi}, \mathbf{y}_{\text{mis}})$ 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 (3.11), $D(\boldsymbol{\psi}) = -2\sum_{i=0}^{n} \Big\{ \mathbf{1}(t_i > 1) \left(\sum_{\ell=1}^{t_i-1} \log(1+\exp(\eta_{i\ell}^{w})) \right) + \mathbf{1}(t_i < T+1) \log(1+\exp(-\eta_{i\ell}^{w})) + \sum_{j=0}^{t_i-2} [r_{ij}\eta_{ij}^{r} - \log(1+\exp(\eta_{ij}^{r}))] \Big\}$. For computing $D(\overline{\boldsymbol{\psi}})$, we need to estimate several discrete parameters such as the count response $\mathbf{y}_{\text{mis}}$. The posterior mean of $\mathbf{y}_{\text{mis}}$, which is no longer count, may not be a desirable estimate to be applied in the $\text{DIC}_{W,\mathbf{R}|\mathbf{y}}$ formula. Instead, we may use the posterior mode, which maintains the count nature of these parameters. Another possible choice given in Huang *et al.* (2005) is that we apply the linear predictor $\eta_{i\ell}^w$, and η_{it}^r directly to the $\text{DIC}_{W,\mathbf{R}|\mathbf{y}}$ formula. Therefore, we have $\text{DIC}_{W,\mathbf{R}|\mathbf{y}} = D(\overline{\boldsymbol{\eta}^w}, \overline{\boldsymbol{\eta}^r}) + 2p_D$, where $\overline{\eta_{i\ell}^w} = E[z_i\alpha_{1t} + \mathbf{x}_{3i}'\boldsymbol{\alpha}_{2t} + h(\boldsymbol{\alpha}_{3t}, \mathbf{y}_{it})|D_{\text{obs}}]$, $\overline{\eta_{it}^r} = E[z_i\phi_{1j} + \mathbf{x}_{4i}'\boldsymbol{\phi}_{2j} + g(\boldsymbol{\phi}_{3j}, \mathbf{R}_{j-1}, \mathbf{y}_j)|D_{\text{obs}}]$, $p_D = \overline{D(\boldsymbol{\psi})} - D(\overline{\boldsymbol{\psi}})$ 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 models for the missing data mechanism depend on $\boldsymbol{\psi}$ only through the linear predictors $\boldsymbol{\eta}^w$ and $\boldsymbol{\eta}^r$. Moreover, with the introduction of $\boldsymbol{\eta}^w$, and $\boldsymbol{\eta}^r$ in the computation of $\text{DIC}_{W,\mathbf{R}|\mathbf{y}}$, we no longer need to worry about the discreteness of the parameters since $\boldsymbol{\eta}^w$ and $\boldsymbol{\eta}^r$ are always continuous. Similar to the traditional DIC, the model with the smallest $\text{DIC}_{W,\mathbf{R}|\mathbf{y}}$ value is the most optimal among all the models under consideration.

$\text{LPML}_{W,\mathbf{R}|\mathbf{y}}$. To assess the missing data mechanism, we adopt the conditional LPML (Hanson *et al.*, 2011), where the pseudomarginal probability, i.e., $\prod_{i=1}^n P(W_i, \mathbf{R}_{it_i}|\mathbf{y}_i, z_i, \mathbf{x}_i, \boldsymbol{\gamma})$, is used to quantify the model's predictive ability. Let $D_{\text{obs}}^{(-i*)} = \{W_j, \mathbf{R}_{jt_j}, j = 1, \ldots, i-1, i+1, \ldots, n\} \cup \{(\mathbf{y}_{j,\text{obs}}, z_j, \mathbf{x}_j), j = 1, \ldots, n\}$ denote the observed data with W_i and $\mathbf{R}_{it_i}$ deleted. Let $\boldsymbol{\psi}_1 = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau, \boldsymbol{\zeta}, \sigma^2, \rho)$, and $\boldsymbol{\psi} = (\boldsymbol{\psi}_1, \boldsymbol{\alpha}, \boldsymbol{\phi})$. Then we have

$$\pi(\boldsymbol{\psi}, \mathbf{y}_{\text{mis}}, \boldsymbol{\epsilon}|D_{\text{obs}}^{(-i*)}) \propto \left\{\prod_{j=1}^n f_y(\mathbf{y}_j|\boldsymbol{\psi}, z_j, \mathbf{x}_j, \boldsymbol{\epsilon}_j) f(\boldsymbol{\epsilon}_j|\sigma^2, \rho)\right\} \times \prod_{j\neq i} f_{W,\mathbf{R}|\mathbf{y}}(W_j, \mathbf{R}_{jt_j}|\boldsymbol{\alpha}, \boldsymbol{\phi}, \mathbf{y}_j, z_j, \mathbf{x}_j)\pi(\boldsymbol{\psi}).$$

The simplified conditional predictive ordinate CPO_i (Chen *et al.*, 2000; Hanson *et al.*, 2011) can be written as

$$\text{CPO}_i = \int \sum_{y_{i,\text{mis}}} f_{W,\mathbf{R}|\mathbf{y}}(W_i, \mathbf{R}_{it_i}|\boldsymbol{\alpha}, \boldsymbol{\phi}, \mathbf{y}_i, z_i, \mathbf{x}_i)\pi(\boldsymbol{\psi}, \mathbf{y}_{\text{mis}}, \boldsymbol{\epsilon}|D_{\text{obs}}^{(-i*)})d\boldsymbol{\epsilon}d\boldsymbol{\psi}$$

$$= \frac{1}{\int \sum_{y_{\text{mis}}} \frac{1}{f_{W,\mathbf{R}|\mathbf{y}}(W_i, \mathbf{R}_{it_i}|\boldsymbol{\alpha}, \boldsymbol{\phi}, \mathbf{y}_i, z_i, \mathbf{x}_i)} \pi(\boldsymbol{\psi}, \mathbf{y}_{\text{mis}}, \boldsymbol{\epsilon}|D_{\text{obs}})d\boldsymbol{\epsilon}d\boldsymbol{\psi}},$$

and the logarithm of the pseudomarginal likelihood is given by

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

Let $\{(\boldsymbol{\psi}_b, \mathbf{y}_{\text{mis},b}, \boldsymbol{\epsilon}_b), b = 1, \ldots, B\}$ denote a Gibbs sample of $(\boldsymbol{\psi}, \mathbf{y}_{\text{mis}}, \boldsymbol{\epsilon})$ from (3.18) and let b represent the b^{th} iteration. A Monte Carlo estimate of CPO_i is given by

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

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

3.3 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 1.3. The baseline covariates in the response model and missing data mechanism include Gender, City, Cohabit, Counselor, Drink, and Age (introduced in Section 2.4). Except for Age, which is continuous, all other covariates are binary. For the missing data mechanism, we also consider covariates $\mathbf{1}(\mathbf{y}_t \geq \mathbf{y}_t^m)$, and $\mathbf{R}_t$ 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.

Table 3.1: Values of $\text{DIC}_{W,\mathbf{R}|\mathbf{y}}$ (p_D) and $\text{LPML}_{W,\mathbf{R}|\mathbf{y}}$ under Ignorable Missingness and Nonignorable Missingness with Various Priors for the HIV Prevention Behavioral Data

Fitted Model	p_D	$\text{DIC}_{W,\mathbf{R}\|\mathbf{y}}$	$\text{LPML}_{W,\mathbf{R}\|\mathbf{y}}$
Ignorable	44.51	4634.79	-2318.85
Nonignorable			
N(0, 1)	117.11	4539.79	-2313.84
N(0, 2)	149.19	4451.91	-2310.67
N(0, 3)	169.03	4396.48	-2309.39
N(0, 4)	166.84	4315.97	-2306.42
N(0, 5)	174.49	4288.92	-2305.63
N(0, 6)	168.71	4265.32	-2304.53
N(0, 7)	186.35	4248.50	-2302.97
N(0, 8)	195.31	4248.02	-2304.15
N(0, 9)	185.49	4224.16	**-2301.43**
N(0, 10)	184.69	**4187.97**	-2302.40
Jeffreys Prior	176.33	4312.23	-2302.26

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.

We fit the ignorable and nonignorable models to the HIV prevention behavioral data. For the ignorable model, we simply set $h(\boldsymbol{\alpha}_{3t}, \mathbf{y}_{it}) = 0$ in (3.5) and $g(\boldsymbol{\phi}_{3j}, \mathbf{R}_{j-1}, \mathbf{y}_j) = 0$ in (3.10). For the nonignorable model, we assumed that $h(\boldsymbol{\alpha}_{3t}, \mathbf{y}_t) = \alpha_{3t1}\mathbf{1}(y_{t-1} \geq y^m_{t-1}) + \alpha_{3t2}\mathbf{1}(y_t \geq y^m_t)$, $g(\boldsymbol{\phi}_{3j}, \mathbf{R}_{j-1}, \mathbf{y}_j) = \boldsymbol{\phi}'_{3j1}\mathbf{R}_{j-1} + \phi_{3j2}\mathbf{1}(y_{j-1} \geq y^m_{j-1}) + \phi_{3j3}\mathbf{1}(y_j \geq y^m_j)$ and considered a $N(0, \sigma^2_{prior})$ prior for α_{3t2}, ϕ_{3j3} as well as Jeffreys prior for $\boldsymbol{\alpha}_t$, $\boldsymbol{\phi}_j$ in (2.14). 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 3.1. As exhibited in Table 3.1, the effective number of parameters under the ignorable model ($p_D = 44.51$) was the smallest among all the models we considered, and approximately equal to the number of parameters of the missing data mechanism. The effective number of parameters under the nonignorable model were significantly larger than the effect number of parameters under the ignorable model.

We also see from Table 3.1 that (i) the DIC value was 4634.79 under the ignorable model; (ii) under the nonignorable model with a $N(0, \sigma^2_{prior})$ prior, the value of DIC tended to decrease as σ^2_{prior} increased; (iii) the DIC attained the local minimum with DIC=4187.97 at $\sigma^2_{prior} = 10$ among all the models under consideration (10 values of σ^2_{prior} and Jeffreys Prior). The results indicated by LPML were different from the results by the DIC criterion. The nonignorable model with a $N(0, 9)$ prior had the largest value of LPML (LPML=-2301.43) among all the models under consideration. These results indicate that for the HIV prevention behavioral data, the missing longitudinal count responses were potentially nonignorably missing.

Tables 3.2-3.4 show the ESTs, SDs, and 95% HPD intervals under the ignorable model, the nonignorable model with the $N(0, 9)$ 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 in Table 3.2, the intervention effects (z) had significant positive posterior estimates (at 12-Month and 18-Month) in the binary model and significant negative posterior estimates (at 6-Month, 12-Month and 18-Month) in the Poisson model.

Gender and Counselor were negatively significant in both Poisson and binary models. Cohabit had significant negative posterior estimates in the binary model while nonsignificant positive posterior estimates in the Poisson model, indicating that people who cohabitated with their primary sex partner were more likely to experience unprotected sex acts. Age had a strong positive effect in the binary model and a strong negative effect in the Poisson model, indicating that older people may have better knowledge of safe sexual behavior. For the missing data mechanism of the intermittent missing, 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 be missing in later period of the study. Females (at 6-Month and 12-Month), people who lived in city (at 12-Month), and people who met with a counselor at least every 3 months (at 12-Month) were less likely to miss their visits. Moreover, people who frequently skipped the previous visits had higher odds of missingness in the future, as indicated by R_0 (at 6-Month) and R_1 (at 12-Month). In addition, people who cohabited with sex partner were more likely to drop out at 6-Month while older people were less likely to drop out at 6-Month. Females were less likely to withdraw at 12-Month and 18-Month.

The posterior estimates in Table 3.3 were similar to those given in Table 3.2. However, City, which is a covariate in the Poisson model, was not significant with 95% HPD interval=(-0.005, 0.333) under the ignorable model but was significant with 95% HPD interval=(0.009, 0.341) under the nonignorable model with a $N(0, 9)$ prior. Similarly, Counselor (at Baseline), which is a covariate in the missing data mechanism (intermittent missing), was not significant with 95% HPD interval=(-0.046, 1.388) in the ignorable case but significant with 95% HPD interval=(0.036, 1.473) in Table 3.3. Moreover, the covariates in the missing data mechanism (intermittent missing), y_0 (95% HPD interval=(0.001, 0.809)) at Baseline, y_0 (95% HPD interval=(-9.040, -1.481)) and y_1 (95% HPD interval=(-9.065, -1.267)) at 6-Month, and y_2 at 12-Month (95% HPD interval=(0.130, 4.083)) were all significant, indicating that missingness of the count responses may be nonignorable. This result was consistent with the DIC and LPML.

Similarly, Cohabit and Age, which are covariates in the missing data mechanism of the withdraw model, were significant with 95% HPD interval=(0.022, 0.859) and (-0.464, -0.008) under the ignorable model but were not significant with 95% HPD interval=(-0.200, 0.728) and (-0.273, 0.349) under the nonignorable model with a $N(0, 9)$ prior.

In addition, the posterior standard deviations in Table 3.3 were slightly smaller than those given in Table 3.2 in the response model. For the covariates in the missing data mechanism shared in both the ignorable and nonignorable models, the posterior standard deviations in Table 3.3 in the missing data mechanism, were similar to those given in Table 3.2. The standard deviation of those variables corresponding to the missing response covariate y_t increased as σ^2_{prior} increased, implying that those variables could not be estimated under an improper uniform prior. It is apparent that the posterior estimates under the nonignorable model were different from those under the ignorable model. The posterior estimates under the nonignorable model with Jeffreys prior (in Table 3.4) were similar to those under the nonignorable model with a $N(0, 9)$ prior (in Table 3.3) for both the response model and the missing data mechanism. The posterior estimates of ρ, σ^2 and τ were similar under the three models.

We note that, it is impossible to directly obtain the overall intervention effects through Table 3.2-3.4 since the intervention effects were contained separately in the binary and

Poisson models. Using the approach developed in Section 3.2.2, we were able to assess the overall intervention effects under the three models, as exhibited in Table 3.5. First of all, the overall intervention effects over time were similar under the three models. Based on the posterior estimates and 95% HPD intervals, the intervention effects varied from negative to positive values as time progressed and were significant at all visits, indicating that the counseling intervention significantly reduced HIV risk behavior immediately after the baseline visit.

Table 3.2: Posterior Summaries under the Ignorable Model for the HIV Prevention Behavioral Data.

	Poisson Response Model		
	EST	SD	95% HPD Interval
Binary			
Intercept	0.774	0.151	(0.475, 1.063)
Gender	-0.594	0.113	(-0.814, -0.370)
City	0.102	0.125	(-0.140, 0.350)
Cohabit	-0.672	0.110	(-0.881, -0.450)
Counselor	-0.729	0.119	(-0.960, -0.495)
Drink	-0.489	0.272	(-1.032, 0.028)
Age	0.291	0.058	(0.178, 0.403)
z			
Baseline	-0.923	0.160	(-1.241, -0.619)
6-Month	0.252	0.155	(-0.059, 0.550)
12-Month	0.737	0.160	(0.425, 1.050)
18-Month	0.716	0.161	(0.404, 1.031)
Poisson			
Intercept	0.350	0.120	(0.119, 0.589)
Gender	-0.302	0.075	(-0.454, -0.159)
City	0.166	0.087	(-0.005, 0.333)
Cohabit	0.127	0.071	(-0.014, 0.263)
Counselor	-0.204	0.094	(-0.388, -0.022)
Drink	0.314	0.177	(-0.043, 0.652)
Age	-0.249	0.039	(-0.327, -0.172)
z			
Baseline	0.289	0.075	(0.134, 0.427)
6-Month	-0.212	0.093	(-0.390, -0.027)
12-Month	-0.363	0.110	(-0.579, -0.151)
18-Month	-0.302	0.112	(-0.522, -0.085)
ρ	0.743	0.026	(0.691, 0.793)
σ^2	0.982	0.027	(0.930, 1.035)
τ	0.453	0.444	(0.000, 1.339)

	Missing Data Mechanism (Intermittent Missing)		
	EST	SD	95% HPD Interval
Baseline			
Intercept	-3.436	0.414	(-4.243, -2.634)
Gender	0.091	0.238	(-0.373, 0.552)
City	-0.306	0.326	(-0.942, 0.333)
Cohabit	0.247	0.228	(-0.185, 0.702)
Counselor	0.668	0.368	(-0.046, 1.388)
Age	0.070	0.112	(-0.149, 0.286)
6-Month			
Intercept	-2.536	0.303	(-3.123, -1.947)
Gender	-0.461	0.192	(-0.851, -0.099)
City	-0.343	0.266	(-0.891, 0.148)
Cohabit	0.022	0.190	(-0.343, 0.401)
Counselor	0.359	0.267	(-0.165, 0.878)
Age	-0.029	0.095	(-0.215, 0.158)
R_0	0.924	0.307	(0.333, 1.525)
12-Month			
Intercept	-2.405	0.303	(-2.993, -1.811)
Gender	-0.518	0.220	(-0.945, -0.089)
City	-0.669	0.332	(-1.338, -0.035)
Cohabit	-0.161	0.215	(-0.591, 0.251)
Counselor	-0.516	0.245	(-0.982, -0.024)
Age	-0.065	0.111	(-0.283, 0.153)
R_0	0.253	0.468	(-0.659, 1.158)
R_1	1.025	0.300	(0.441, 1.608)
z			
Baseline	-0.636	0.232	(-1.070, -0.167)
6-Month	-0.113	0.186	(-0.479, 0.250)
12-Month	0.667	0.223	(0.239, 1.107)

	Missing Data Mechanism (Dropout)		
	EST	SD	95% HPD Interval
6-Month			
Intercept	-3.203	0.333	(-3.864, -2.567)
Gender	-0.337	0.218	(-0.760, 0.089)
City	0.423	0.243	(-0.059, 0.892)
Cohabit	0.431	0.213	(0.022, 0.859)
Counselor	0.142	0.279	(-0.407, 0.683)
Age	-0.230	0.117	(-0.464, -0.008)
18-Month			
Intercept	-2.542	0.281	(-3.106, -2.008)
Gender	-0.476	0.193	(-0.847, -0.091)
City	0.349	0.222	(-0.112, 0.760)
Cohabit	0.188	0.187	(-0.178, 0.553)
Counselor	0.070	0.242	(-0.382, 0.564)

	Missing Data Mechanism (Dropout)		
	EST	SD	95% HPD Interval
12-Month			
Intercept	-3.048	0.380	(-3.802, -2.330)
Gender	-0.667	0.257	(-1.179, -0.175)
City	-0.241	0.341	(-0.926, 0.406)
Cohabit	-0.355	0.254	(-0.834, 0.158)
Counselor	0.236	0.333	(-0.408, 0.886)
Age	-0.208	0.133	(-0.472, 0.050)
z			
6-Month	0.014	0.208	(-0.397, 0.413)
12-Month	0.273	0.251	(-0.216, 0.771)
18-Month	0.055	0.186	(-0.309, 0.418)

	EST	SD	95% HPD Interval
Age	-0.062	0.098	(-0.250, 0.134)

Table 3.3: Posterior Summaries under the Nonignorable Model with a $N(0,9)$ Prior for the HIV Prevention Behavioral Data.

	Poisson Response Model		
	EST	SD	95% HPD Interval
Binary			
Intercept	0.744	0.143	(0.466, 1.029)
Gender	-0.575	0.108	(-0.785, -0.361)
City	0.064	0.120	(-0.182, 0.291)
Cohabit	-0.675	0.105	(-0.878, -0.467)
Counselor	-0.703	0.114	(-0.925, -0.475)
Drink	-0.408	0.251	(-0.910, 0.070)
Age	0.284	0.055	(0.177, 0.394)
z			
Baseline	-0.817	0.155	(-1.127, -0.520)
6-Month	0.226	0.151	(-0.070, 0.520)
12-Month	0.680	0.157	(0.368, 0.980)
18-Month	0.648	0.157	(0.339, 0.954)
Poisson			
Intercept	0.450	0.118	(0.215, 0.681)
Gender	-0.336	0.074	(-0.478, -0.188)
City	0.174	0.085	(0.009, 0.341)
Cohabit	0.107	0.070	(-0.028, 0.244)
Counselor	-0.234	0.091	(-0.416, -0.062)
Drink	0.300	0.171	(-0.028, 0.639)
Age	-0.244	0.039	(-0.321, -0.167)
z			
Baseline	0.245	0.074	(0.100, 0.390)
6-Month	-0.173	0.092	(-0.358, 0.005)
12-Month	-0.289	0.109	(-0.505, -0.079)
18-Month	-0.200	0.124	(-0.441, 0.040)
ρ	0.738	0.028	(0.682, 0.790)
σ^2	0.985	0.027	(0.931, 1.037)
τ	0.438	0.425	(0.000, 1.272)

	Missing Data Mechanism (Intermittent Missing)		
	EST	SD	95% HPD Interval
Baseline			
Intercept	-3.061	0.408	(-3.849, -2.258)
Gender	0.051	0.235	(-0.403, 0.512)
City	-0.207	0.329	(-0.866, 0.424)
Cohabit	0.324	0.230	(-0.127, 0.772)
Counselor	0.741	0.368	(0.036, 1.473)
Age	-0.021	0.116	(-0.249, 0.206)
y_0	-3.987	2.016	(-9.040, -1.481)
6-Month			
Intercept	-2.412	0.314	(-3.028, -1.799)
Gender	-0.538	0.197	(-0.925, -0.150)
City	-0.350	0.271	(-0.893, 0.160)
Cohabit	0.057	0.192	(-0.317, 0.433)
Counselor	0.403	0.273	(-0.127, 0.941)
Age	-0.116	0.101	(-0.322, 0.075)
R_0	1.202	0.327	(0.535, 1.811)
y_0	0.408	0.205	(0.001, 0.809)
y_1	-3.971	2.037	(-9.065, -1.267)
12-Month			
Intercept	-3.047	0.507	(-4.041, -2.075)
Gender	-0.654	0.246	(-1.124, -0.156)
City	-0.702	0.362	(-1.449, -0.030)
Cohabit	-0.318	0.242	(-0.808, 0.138)
Counselor	-0.593	0.270	(-1.129, -0.067)
Age	0.047	0.128	(-0.202, 0.297)
R_0	0.393	0.503	(-0.588, 1.396)
R_1	0.883	0.348	(0.194, 1.558)
y_1	-0.672	0.452	(-1.546, 0.283)
y_2	2.130	1.151	(0.130, 4.083)
z			
Baseline			
6-Month	-0.615	0.236	(-1.071, -0.146)
12-Month	-0.206	0.187	(-0.572, 0.158)
18-Month	0.889	0.262	(0.378, 1.394)

	Missing Data Mechanism (Dropout)		
	EST	SD	95% HPD Interval
6-Month			
Intercept	-5.197	1.472	(-8.101, -2.765)
Gender	-0.202	0.238	(-0.681, 0.245)
City	0.455	0.264	(-0.057, 0.979)

	Missing Data Mechanism (Dropout)		
	EST	SD	95% HPD Interval
12-Month			
Intercept	-3.035	0.627	(-3.863, -2.134)
Gender	-0.648	0.266	(-1.156, -0.117)
City	-0.266	0.344	(-0.962, 0.379)

Cohabit	0.269	0.239	(-0.200, 0.728)	Cohabit	-0.283	0.263	(-0.789, 0.241)
Counselor	-0.019	0.312	(-0.612, 0.608)	Counselor	0.229	0.339	(-0.423, 0.900)
Age	0.047	0.158	(-0.273, 0.349)	Age	-0.220	0.141	(-0.496, 0.058)
y_0	-0.765	0.360	(-1.452, 0.001)	y_1	0.426	0.444	(-0.380, 1.233)
y_1	3.716	1.940	(-0.309, 7.584)	y_2	-3.021	2.380	(-8.541, 0.477)
18-Month				**z**			
Intercept	-3.174	0.763	(-4.371, -2.063)	quad 6-Month	0.223	0.240	(-0.240, 0.694)
Gender	-0.553	0.213	(-0.966, -0.132)	12-Month	0.231	0.262	(-0.259, 0.772)
City	0.445	0.249	(-0.040, 0.932)	18-Month	0.144	0.206	(-0.256, 0.547)
Cohabit	0.117	0.206	(-0.282, 0.525)				
Counselor	-0.132	0.287	(-0.721, 0.405)				
Age	0.050	0.121	(-0.186, 0.289)				
y_2	-0.809	0.427	(-1.656, 0.043)				
y_3	2.185	1.329	(-0.518, 4.709)				

Table 3.4: Posterior Summaries under the Nonignorable Model with Jeffreys Prior for the HIV Prevention Behavioral Data.

	Poisson Response Model				Missing Data Mechanism (Intermittent Missing)		
	EST	SD	95% HPD Interval		EST	SD	95% HPD Interval
Binary				**Baseline**			
Intercept	0.743	0.145	(0.466, 1.040)	Intercept	-2.984	0.406	(-3.776, -2.200)
Gender	-0.576	0.109	(-0.791, -0.365)	Gender	0.049	0.233	(-0.410, 0.502)
City	0.070	0.122	(-0.174, 0.304)	City	-0.172	0.320	(-0.804, 0.451)
Cohabit	-0.676	0.106	(-0.879, -0.464)	Cohabit	0.324	0.231	(-0.123, 0.774)
Counselor	-0.704	0.116	(-0.926, -0.474)	Counselor	0.688	0.364	(0.007, 1.423)
Drink	-0.405	0.253	(-0.906, 0.084)	Age	-0.020	0.116	(-0.255, 0.199)
Age	0.284	0.056	(0.176, 0.396)	y_0	-3.965	2.138	(-9.106, -1.224)
				6-Month			
z				Intercept	-2.374	0.309	(-2.983, -1.777)
Baseline	-0.832	0.156	(-1.132, -0.519)	Gender	-0.529	0.195	(-0.902, -0.138)
6-Month	0.236	0.153	(-0.058, 0.539)	City	-0.325	0.268	(-0.844, 0.202)
12-Month	0.687	0.158	(0.377, 0.994)	Cohabit	0.061	0.189	(-0.311, 0.430)
18-Month	0.663	0.156	(0.359, 0.967)	Counselor	0.384	0.270	(-0.146, 0.904)
				Age	-0.110	0.100	(-0.297, 0.095)
Poisson				R_0	1.160	0.329	(0.521, 1.798)
Intercept	0.427	0.118	(0.202, 0.666)	y_0	0.403	0.206	(0.002, 0.810)
Gender	-0.327	0.075	(-0.473, -0.179)	y_1	-3.975	2.252	(-9.075, -0.966)
City	0.172	0.086	(0.004, 0.341)				
Cohabit	0.110	0.070	(-0.028, 0.245)	**12-Month**			
Counselor	-0.229	0.093	(-0.413, -0.049)	Intercept	-2.882	0.442	(-3.749, -2.027)
Drink	0.309	0.172	(-0.034, 0.638)	Gender	-0.625	0.239	(-1.094, -0.155)
Age	-0.248	0.040	(-0.324, -0.170)	City	-0.641	0.346	(-1.324, 0.038)
				Cohabit	-0.301	0.234	(-0.761, 0.153)
z				Counselor	-0.583	0.260	(-1.085, -0.066)
Baseline	0.252	0.074	(0.101, 0.390)	Age	0.037	0.123	(-0.203, 0.279)
6-Month	-0.180	0.094	(-0.367, 0.001)	R_0	0.441	0.485	(-0.495, 1.413)
12-Month	-0.298	0.113	(-0.522, -0.081)	R_1	0.853	0.335	(0.181, 1.501)
18-Month	-0.217	0.117	(-0.439, 0.011)	y_1	-0.580	0.431	(-1.409, 0.304)
				y_2	1.904	0.974	(0.115, 3.582)
ρ	0.739	0.027	(0.686, 0.789)				

	EST	SD	95% HPD Interval		EST	SD	95% HPD Interval
σ^2	0.984	0.027	(0.931, 1.036)	z			
τ	0.440	0.426	(0.000, 1.293)	Baseline			
				6-Month	-0.605	0.234	(-1.045, -0.132)
				12-Month	-0.205	0.188	(-0.580, 0.157)
				18-Month	0.845	0.249	(0.346, 1.320)

	Missing Data Mechanism (Dropout)				Missing Data Mechanism (Dropout)		
	EST	SD	95% HPD Interval		EST	SD	95% HPD Interval
6-Month				**12-Month**			
Intercept	-4.612	1.354	(-7.414, -2.560)	Intercept	-2.928	0.457	(-3.757, -2.120)
Gender	-0.218	0.231	(-0.673, 0.230)	Gender	-0.632	0.257	(-1.136, -0.137)
City	0.455	0.257	(-0.051, 0.953)	City	-0.229	0.336	(-0.883, 0.432)
Cohabit	0.280	0.237	(-0.196, 0.740)	Cohabit	-0.282	0.261	(-0.795, 0.224)
Counselor	-0.024	0.302	(-0.603, 0.570)	Counselor	0.188	0.326	(-0.436, 0.836)
Age	0.014	0.162	(-0.318, 0.318)	Age	-0.210	0.139	(-0.473, 0.070)
y_0	-0.646	0.380	(-1.333, 0.172)	y_1	0.438	0.425	(-0.408, 1.205)
y_1	2.997	1.900	(-0.954, 6.971)	y_2	-2.862	2.407	(-8.399, 0.952)
18-Month				**z**			
Intercept	-2.915	0.449	(-3.812, -2.079)	6-Month	0.194	0.235	(-0.273, 0.650)
Gender	-0.527	0.205	(-0.921, -0.111)	12-Month	0.221	0.257	(-0.291, 0.718)
City	0.432	0.238	(-0.044, 0.886)	18-Month	0.132	0.199	(-0.241, 0.536)
Cohabit	0.126	0.201	(-0.277, 0.505)				
Counselor	-0.112	0.272	(-0.644, 0.418)				
Age	0.033	0.115	(-0.195, 0.254)				
y_2	-0.674	0.420	(-1.439, 0.213)				
y_3	1.736	1.138	(-0.185, 3.849)				

Table 3.5: Posterior Summaries of the Adjusted Overall Intervention Effect under the Ignorable Model, Nonignorable Model with a N(0, 9) Prior and Jeffreys Prior.

	Ignorable		
	EST	SD	95% HPD Interval
z			
Baseline	0.504	0.067	(0.373, 0.636)
6-Month	-0.351	0.081	(-0.504, -0.187)
12-Month	-0.727	0.095	(-0.922, -0.549)
18-Month	-0.655	0.095	(-0.839, -0.469)
	Nonignorable N(0, 9) Prior		
	EST	SD	95% HPD Interval
z			
Baseline	0.445	0.068	(0.308, 0.573)
6-Month	-0.296	0.087	(-0.468, -0.127)
12-Month	-0.620	0.105	(-0.828, -0.416)
18-Month	-0.516	0.122	(-0.750, -0.275)
	Nonignorable Jeffreys Prior		
	EST	SD	95% HPD Interval
z			
Baseline	0.453	0.069	(0.322, 0.591)
6-Month	-0.308	0.089	(-0.483, -0.133)
12-Month	-0.633	0.106	(-0.844, -0.431)
18-Month	-0.541	0.113	(-0.767, -0.324)

Chapter 4

Assessment of Missing Data Mechanism

4.1 Motivation

Two different missing data mechanisms are introduced in Chapter 2 and Chapter 3. In Chapter 2, we assume a logistic regression model for modeling $P(R_t = 1|\mathbf{R}_{t-1}, \mathbf{y}_t, z, \mathbf{x}_2, \boldsymbol{\gamma}_t)$, and construct the joint distribution of $\mathbf{R}$ via a sequence of one-dimensional conditional distributions. This conventional model is easy to implement, but does not inherently take into account the difference between intermittent missing and dropout. Instead, the missing data mechanism proposed in Chapter 3 has the nice property of directly modeling dropout and intermittent missing. To be more specific, we first model prospectively when each subject will drop out. Given that the subject withdraws at a certain visit, a joint model of all the previous missing status is then developed. This model is especially appropriate for some clinical trial studies where patients notify the investigator beforehand if they decide to drop out at a certain visit. The model also aligns well with particular types of experimental designs, in which patients automatically leave from the longitudinal study once their diseases are cured. One motivating example is the pain sensitivity study of low back pain (Starkweather *et al.* (2016)). The sample consisted of 48 participants, of whom 19 went on to develop persistent low back pain and 29 resolved within the first 6 weeks after initial onset. Those who resolved midway during the trial automatically withdrew from the study the moment their low pack pains were cured. The prequential multinomial model in (3.4) is thus a natural and perfect fit for this problem.

Another interesting finding of the HIV prevention behavioral data is that the longitudinal binary response variable considered in Chapter 2 is actually the dichotomization of the longitudinal count data in Chapter 3. To be more specific, the binary response variable takes value of zero if and only if the corresponding count response variable, i.e., the total number of ACASI-reported unprotected sex acts in the past 4 weeks with partners of any HIV status, equals zero. In other word, the two longitudinal response variables are actually measuring the same thing. However, the longitudinal count response variable carries more information than the binary response variable. Therefore, we expect the performance of the missing data mechanism corresponding to the longitudinal count response variable to be better since the data contain more information. Note that the joint DIC and LPML criteria are no longer suitable to assess the performances of the missing data mechanism, since the response variables are now different. Therefore, we still apply the Bayesian model assessment criteria, namely, the DIC relating to the missing data model, and the LMPL relating to the missing data model developed in Section 2.2.4 and Section 3.2.5.

4.2 Assessment of Missing Data Mechanism in Section 2.1.2

We assume the zero-inflated Poisson distribution in Section 3.1.1 for the response variable, and the model developed in Section 2.1.2 for the missing data mechanism. We then fit both ignorable and nonignorable models to the count response data. For the

Table 4.1: Values of $\text{DIC}_{\mathbf{R}|\mathbf{y}}$ (p_D) and $\text{LPML}_{\mathbf{R}|\mathbf{y}}$ for Longitudinal Binary Variable and Longitudinal Poisson Variable under Ignorable Missingness and Nonignorable Missingness with Various Priors for the HIV Prevention Behavioral Data

	Binary			Poisson		
Fitted Model	p_D	$\text{DIC}_{\mathbf{R}\|\mathbf{y}}$	$\text{LPML}_{\mathbf{R}\|\mathbf{y}}$	p_D	$\text{DIC}_{\mathbf{R}\|\mathbf{y}}$	$\text{LPML}_{\mathbf{R}\|\mathbf{y}}$
Ignorable	30.85	4793.16	-2397.24	30.95	4793.36	-2397.34
Nonignorable						
N(0, 1)	89.82	4769.73	-2398.26	75.87	4718.05	-2392.04
N(0, 2)	107.06	4755.71	-2397.44	84.70	4691.24	-2390.83
N(0, 3)	114.95	4757.82	-2397.86	87.68	4678.77	-2390.16
N(0, 4)	112.99	4751.86	-2397.70	89.07	4671.40	-2390.30
N(0, 5)	126.66	4748.78	-2397.28	90.15	4662.70	-2389.46
N(0, 6)	132.95	4746.74	-2397.23	91.47	4658.82	-2389.67
N(0, 7)	132.67	4747.22	-2397.23	93.47	4654.31	**-2388.25**
N(0, 8)	132.94	**4737.61**	**-2396.32**	91.55	4650.99	-2389.10
N(0, 9)	133.47	4745.62	-2397.29	93.60	4647.21	-2388.29
N(0, 10)	140.61	4749.97	-2398.21	92.92	**4643.92**	-2389.23
Jeffreys Prior	120.18	4750.08	**-2396.64**	94.30	4662.50	**-2386.80**

ignorable model, we simply set $h(\mathbf{y}_t, \boldsymbol{\gamma}_{4t}) = 0$ in (2.8). For the nonignorable model, we assumed that $h(\mathbf{y}_t, \boldsymbol{\gamma}_{4t}) = \gamma_{4t1} y_{t-1} + \gamma_{4t2} y_t$ in (2.8) and considered a $N(0, \sigma^2_{prior})$ prior for γ_{4t2} as well as Jeffreys prior for $\boldsymbol{\gamma}_t$ in (2.14). We specified uniform priors for all other parameters. Similar to Section 2.4 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 of the count data are shown on the last three columns in Table 4.1. The effective number of parameters under the ignorable model ($p_D = 30.95$) was the smallest among all the models we considered, and approximately equal to the number of parameters of the missing data mechanism. Under the nonignorable model with a $N(0, \sigma^2_{prior})$ prior, the effective number of parameters increased with σ^2_{prior}. We also see from Table 4.1 that (i) the DIC value was 4793.36 under the ignorable model; (ii) under the nonignorable model with a $N(0, \sigma^2_{prior})$ prior, the value of DIC steady decreased as σ^2_{prior}

increased; (iii) the DIC attained the local minimum with DIC=4643.92 at $\sigma^2_{prior} = 10$ among all the models under consideration (10 values of σ^2_{prior} and Jeffreys Prior).

The results indicated by LPML were slightly different from the results by the DIC criterion. The nonignorable model with Jeffreys prior had the largest value of LPML (LPML=-2386.80) among all the models under consideration. The nonignorable model with a $N(0, 7)$ prior had the second largest value of LPML (LPML=-2388.25). These results indicate that for the HIV prevention behavioral data, the missing longitudinal count responses were potentially nonignorably missing.

We further see from Table 4.1 that (i) under the ignorable model, the DIC (4793.16) and LPML (-2397.24) values of the binary response variable were similar to the DIC (4793.36) and LPML (-2397.34) values of the count response variable; (ii) under the non-ignorable models, all of the DIC values of the count response variable were smaller than those of the binary response variable, indicating that the missing data mechanism of the count response variable performed better; (iii) under the nonignorable models, all of the LPML values of the count response variable were larger than those of the binary response variable, again, implying that the missing data mechanism of the count response variable performed better. These findings confirmed our conjecture that the count response variable contained more information than the binary response variable and therefore should improve the fit of the missing data mechanism.

As exhibited in Table 4.2, the overall intervention effects over time were similar under the three models (ignorable, nonignorable with a $N(0, 7)$ Prior and Jeffreys Prior). Based on the posterior estimates and 95% HPD intervals, the intervention effects varied from negative to positive values as time progressed and were significant at all visits, indicating

Table 4.2: Posterior Summaries of the Adjusted Overall Intervention Effect under the Ignorable Model, Nonignorable Model with a N(0, 7) Prior and Jeffreys Prior.

	Ignorable		
	EST	SD	95% HPD Interval
z			
Baseline	0.505	0.067	(0.373, 0.636)
6-Month	-0.351	0.083	(-0.517, -0.192)
12-Month	-0.728	0.094	(-0.924, -0.555)
18-Month	-0.659	0.096	(-0.847, -0.471)
	Nonignorable N(0, 7) Prior		
	EST	SD	95% HPD Interval
z			
Baseline	0.493	0.069	(0.360, 0.629)
6-Month	-0.408	0.083	(-0.570, -0.245)
12-Month	-0.580	0.117	(-0.809, -0.352)
18-Month	-0.640	0.098	(-0.837, -0.451)
	Nonignorable Jeffreys Prior		
	EST	SD	95% HPD Interval
z			
Baseline	0.491	0.069	(0.355, 0.626)
6-Month	-0.406	0.084	(-0.572, -0.243)
12-Month	-0.584	0.111	(-0.796, -0.364)
18-Month	-0.637	0.100	(-0.839, -0.446)

that the counseling intervention significantly reduced HIV risk behavior immediately after the baseline visit.

4.3 Assessment of Missing Data Mechanism in Section 3.1.2

We assume the probit mixed-effects regression model in Section 2.1.1 for the response variable, and the models developed in Section 3.1.2 for the missing data mechanism. We then fit both ignorable and nonignorable models to the count response data. For the ignorable model, we simply set $h(\boldsymbol{\alpha}_{3t}, \mathbf{y}_{it}) = 0$ in (3.5) and $g(\boldsymbol{\phi}_{3j}, \mathbf{R}_{j-1}, \mathbf{y}_j) = 0$ in (3.10). For the nonignorable model, we assumed that $h(\boldsymbol{\alpha}_{3t}, \mathbf{y}_t) = \alpha_{3t1}\mathbf{1}(y_{t-1} \geq y^m_{t-1}) + \alpha_{3t2}\mathbf{1}(y_t \geq y^m_t)$, $g(\boldsymbol{\phi}_{3j}, \mathbf{R}_{j-1}, \mathbf{y}_j) = \boldsymbol{\phi}'_{3j1}\mathbf{R}_{j-1} + \phi_{3j2}\mathbf{1}(y_{j-1} \geq y^m_{j-1}) + \phi_{3j3}\mathbf{1}(y_j \geq y^m_j)$ and considered a $N(0, \sigma^2_{prior})$ prior for α_{3t2}, ϕ_{3j3} as well as Jeffreys prior for $\boldsymbol{\alpha}_t$, $\boldsymbol{\phi}_j$ in (2.14). We specified uniform priors for all other parameters. Similar to Section 3.3 we then computed DIC and

Table 4.3: Posterior Summaries under the Ignorable Model for the HIV Prevention Behavioral Data.

	Poisson Response Model		
	EST	SD	95% HPD Interval
Binary			
Intercept	0.772	0.148	(0.480, 1.050)
Gender	-0.587	0.111	(-0.810, -0.374)
City	0.104	0.125	(-0.142, 0.346)
Cohabit	-0.673	0.113	(-0.893, -0.449)
Counselor	-0.730	0.118	(-0.963, -0.501)
Drink	-0.475	0.266	(-1.031, 0.010)
Age	0.290	0.058	(0.175, 0.403)
z			
Baseline	-0.923	0.160	(-1.238, -0.616)
6-Month	0.253	0.155	(-0.053, 0.550)
12-Month	0.734	0.161	(0.426, 1.053)
18-Month	0.707	0.163	(0.383, 1.019)
Poisson			
Intercept	0.351	0.117	(0.116, 0.572)
Gender	-0.298	0.075	(-0.448, -0.155)
City	0.168	0.086	(0.002, 0.337)
Cohabit	0.125	0.072	(-0.014, 0.264)
Counselor	-0.207	0.092	(-0.390, -0.026)
Drink	0.318	0.177	(-0.037, 0.658)
Age	-0.249	0.039	(-0.328, -0.174)
z			
Baseline	0.289	0.075	(0.141, 0.435)
6-Month	-0.212	0.094	(-0.397, -0.028)
12-Month	-0.365	0.111	(-0.574, -0.144)
18-Month	-0.310	0.114	(-0.527, -0.082)
ρ	0.743	0.027	(0.688, 0.793)
σ^2	0.982	0.027	(0.929, 1.034)
τ	0.457	0.449	(0.000, 1.347)

	Missing Data Mechanism		
	EST	SD	95% HPD Interval
Baseline			
Intercept	-3.490	0.418	(-4.337, -2.714)
Gender	0.113	0.236	(-0.347, 0.574)
City	-0.337	0.326	(-0.987, 0.294)
Cohabit	0.225	0.224	(-0.215, 0.659)
Counselor	0.666	0.371	(-0.028, 1.417)
Age	0.083	0.112	(-0.132, 0.305)
6-Month			
Intercept	-2.095	0.231	(-2.567, -1.659)
Gender	-0.397	0.147	(-0.681, -0.105)
City	0.032	0.183	(-0.337, 0.377)
Cohabit	0.220	0.145	(-0.061, 0.502)
Counselor	0.269	0.197	(-0.104, 0.666)
Age	-0.102	0.076	(-0.254, 0.043)
R_0	0.364	0.300	(-0.238, 0.946)
12-Month			
Intercept	-1.949	0.212	(-2.365, -1.538)
Gender	-0.483	0.145	(-0.761, -0.191)
City	-0.117	0.183	(-0.470, 0.247)
Cohabit	-0.108	0.141	(-0.386, 0.162)
Counselor	-0.250	0.174	(-0.590, 0.085)
Age	-0.160	0.074	(-0.307, -0.014)
$\sum_{j=0}^{1} R_j$	1.643	0.141	(1.376, 1.924)
18-Month			
Intercept	-2.638	0.240	(-3.122, -2.185)
Gender	-0.382	0.153	(-0.682, -0.084)
City	0.400	0.185	(0.027, 0.751)
Cohabit	0.079	0.150	(-0.213, 0.378)
Counselor	0.074	0.195	(-0.299, 0.460)
Age	-0.127	0.078	(-0.279, 0.027)
$\sum_{j=0}^{1} R_j$	1.775	0.103	(1.580, 1.982)
z			
Baseline	-0.634	0.231	(-1.111, -0.199)
6-Month	-0.073	0.142	(-0.351, 0.205)
12-Month	0.454	0.144	(0.181, 0.743)
18-Month	0.134	0.148	(-0.161, 0.418)

Table 4.4: Posterior Summaries under the Nonignorable Model with a $N(0,7)$ Prior for the HIV Prevention Behavioral Data.

	Poisson Response Model				Missing Data Mechanism		
	EST	SD	95% HPD Interval		EST	SD	95% HPD Interval
Binary				**Baseline**			
Intercept	0.757	0.148	(0.453, 1.042)	Intercept	-3.119	0.405	(-3.909, -2.319)
Gender	-0.572	0.112	(-0.788, -0.345)	Gender	0.082	0.234	(-0.390, 0.531)
City	0.100	0.122	(-0.147, 0.332)	City	-0.249	0.334	(-0.919, 0.388)
Cohabit	-0.647	0.110	(-0.867, -0.437)	Cohabit	0.314	0.230	(-0.153, 0.748)
Counselor	-0.701	0.118	(-0.937, -0.472)	Counselor	0.732	0.368	(0.053, 1.487)
Drink	-0.415	0.260	(-0.936, 0.084)	Age	-0.003	0.114	(-0.225, 0.222)
Age	0.280	0.058	(0.169, 0.396)	y_0	-3.967	1.840	(-8.824, -1.615)
z				**6-Month**			
Baseline	-0.902	0.161	(-1.216, -0.589)	Intercept	-1.998	0.238	(-2.466, -1.534)
6-Month	0.332	0.155	(0.032, 0.638)	Gender	-0.456	0.152	(-0.758, -0.163)
12-Month	0.587	0.164	(0.262, 0.902)	City	0.032	0.187	(-0.327, 0.400)
18-Month	0.699	0.163	(0.377, 1.014)	Cohabit	0.252	0.147	(-0.030, 0.543)
				Counselor	0.321	0.201	(-0.079, 0.704)
Poisson				Age	-0.171	0.082	(-0.335, -0.014)
Intercept	0.353	0.118	(0.124, 0.585)	R_0	0.565	0.319	(-0.071, 1.176)
Gender	-0.299	0.076	(-0.447, -0.146)	y_0	0.298	0.180	(-0.057, 0.650)
City	0.164	0.086	(-0.002, 0.338)	y_1	-2.460	1.805	(-6.145, 0.254)
Cohabit	0.119	0.071	(-0.016, 0.261)				
Counselor	-0.207	0.092	(-0.386, -0.023)	**12-Month**			
Drink	0.308	0.176	(-0.027, 0.660)	Intercept	-2.139	0.273	(-2.662, -1.598)
Age	-0.251	0.039	(-0.327, -0.175)	Gender	-0.534	0.153	(-0.843, -0.241)
				City	-0.110	0.191	(-0.477, 0.272)
z				Cohabit	-0.183	0.155	(-0.486, 0.119)
Baseline	0.274	0.076	(0.125, 0.419)	Counselor	-0.260	0.183	(-0.618, 0.095)
6-Month	-0.235	0.093	(-0.415, -0.052)	Age	-0.121	0.084	(-0.283, 0.044)
12-Month	-0.287	0.116	(-0.512, -0.053)	$\sum_{j=0}^{1} R_j$	1.619	0.153	(1.315, 1.917)
18-Month	-0.291	0.114	(-0.514, -0.069)	y_1	-0.652	0.338	(-1.329, -0.003)
				y_2	1.148	0.664	(-0.191, 2.343)
ρ	0.738	0.027	(0.684, 0.791)				
σ^2	0.982	0.027	(0.931, 1.037)	**18-Month**			
τ	0.461	0.451	(0.000, 1.346)	Intercept	-2.701	0.248	(-3.191, -2.227)
				Gender	-0.384	0.153	(-0.679, -0.076)
				City	0.410	0.184	(0.050, 0.768)
				Cohabit	0.057	0.152	(-0.239, 0.354)
				Counselor	0.073	0.196	(-0.309, 0.457)
				Age	-0.110	0.080	(-0.268, 0.045)
				$\sum_{j=0}^{1} R_j$	1.760	0.106	(1.561, 1.974)
				$\mathbf{y}_2$	0.317	0.300	(-0.292, 0.880)
				$\mathbf{y}_3$	-0.119	0.454	(-1.011, 0.778)
				z			
				Baseline	-0.604	0.234	(-1.070, -0.154)
				6-Month	-0.137	0.148	(-0.416, 0.161)
				12-Month	0.547	0.163	(0.226, 0.863)
				18-Month	0.152	0.151	(-0.143, 0.449)

Table 4.5: Posterior Summaries under the Nonignorable Model with Jeffreys Prior for the HIV Prevention Behavioral Data.

	Poisson Response Model		
	EST	SD	95% HPD Interval
Binary			
Intercept	0.762	0.148	(0.470, 1.049)
Gender	-0.574	0.110	(-0.784, -0.353)
City	0.096	0.124	(-0.147, 0.339)
Cohabit	-0.651	0.109	(-0.864, -0.442)
Counselor	-0.705	0.120	(-0.937, -0.464)
Drink	-0.415	0.263	(-0.929, 0.101)
Age	0.282	0.057	(0.168, 0.391)
z			
Baseline	-0.902	0.160	(-1.230, -0.601)
6-Month	0.329	0.150	(0.033, 0.620)
12-Month	0.597	0.165	(0.272, 0.920)
18-Month	0.700	0.159	(0.392, 1.016)
Poisson			
Intercept	0.359	0.119	(0.126, 0.590)
Gender	-0.300	0.075	(-0.443, -0.149)
City	0.163	0.087	(-0.004, 0.336)
Cohabit	0.117	0.070	(-0.017, 0.257)
Counselor	-0.211	0.093	(-0.394, -0.029)
Drink	0.309	0.176	(-0.035, 0.649)
Age	-0.250	0.039	(-0.328, -0.176)
z			
Baseline	0.272	0.076	(0.127, 0.424)
6-Month	-0.235	0.093	(-0.417, -0.053)
12-Month	-0.287	0.109	(-0.501, -0.072)
18-Month	-0.288	0.113	(-0.511, -0.072)
ρ	0.739	0.027	(0.685, 0.790)
σ^2	0.982	0.027	(0.928, 1.033)
τ	0.456	0.453	(0.000, 1.340)

	Missing Data Mechanism		
	EST	SD	95% HPD Interval
Baseline			
Intercept	-3.028	0.403	(-3.823, -2.251)
Gender	0.079	0.233	(-0.373, 0.534)
City	-0.211	0.323	(-0.853, 0.412)
Cohabit	0.313	0.230	(-0.149, 0.754)
Counselor	0.675	0.363	(0.005, 1.421)
Age	-0.002	0.114	(-0.223, 0.221)
y_0	-4.456	2.234	(-9.237, -1.567)
6-Month			
Intercept	-1.972	0.239	(-2.464, -1.529)
Gender	-0.451	0.153	(-0.742, -0.144)
City	0.037	0.184	(-0.333, 0.391)
Cohabit	0.247	0.147	(-0.044, 0.530)
Counselor	0.304	0.202	(-0.091, 0.697)
Age	-0.166	0.082	(-0.326, -0.007)
R_0	0.538	0.319	(-0.073, 1.170)
y_0	0.289	0.181	(-0.073, 0.638)
y_1	-2.446	2.099	(-7.311, 0.576)
12-Month			
Intercept	-2.102	0.258	(-2.614, -1.606)
Gender	-0.526	0.152	(-0.823, -0.230)
City	-0.102	0.189	(-0.470, 0.264)
Cohabit	-0.179	0.152	(-0.477, 0.117)
Counselor	-0.263	0.179	(-0.616, 0.082)
Age	-0.119	0.082	(-0.272, 0.049)
$\sum_{j=0}^{1} R_j$	1.602	0.152	(1.301, 1.895)
y_1	-0.620	0.321	(-1.247, 0.013)
y_2	1.124	0.602	(-0.083, 2.206)
18-Month			
Intercept	-2.667	0.248	(-3.157, -2.192)
Gender	-0.381	0.153	(-0.688, -0.086)
City	0.414	0.182	(0.056, 0.773)
Cohabit	0.055	0.150	(-0.235, 0.348)
Counselor	0.060	0.198	(-0.326, 0.451)
Age	-0.106	0.080	(-0.269, 0.043)
$\sum_{j=0}^{1} R_j$	1.743	0.105	(1.547, 1.957)
$\mathbf{y}_2$	0.309	0.290	(-0.251, 0.877)
$\mathbf{y}_3$	-0.077	0.439	(-0.911, 0.821)
z			
Baseline	-0.596	0.232	(-1.065, -0.156)
6-Month	-0.135	0.146	(-0.431, 0.142)
12-Month	0.534	0.156	(0.231, 0.840)
18-Month	0.154	0.152	(-0.148, 0.451)

Table 4.6: Values of $\text{DIC}_{W,\mathbf{R}|\mathbf{y}}$ (p_D) and $\text{LPML}_{W,\mathbf{R}|\mathbf{y}}$ for Longitudinal Binary Variable and Longitudinal Poisson Variable under Ignorable Missingness and Nonignorable Missingness with Various Priors for the HIV Prevention Behavioral Data

	Binary			Poisson		
Fitted Model	p_D	$\text{DIC}_{W,\mathbf{R}\mid\mathbf{y}}$	$\text{LPML}_{W,\mathbf{R}\mid\mathbf{y}}$	p_D	$\text{DIC}_{W,\mathbf{R}\mid\mathbf{y}}$	$\text{LPML}_{W,\mathbf{R}\mid\mathbf{y}}$
Ignorable	44.58	4634.93	-2318.96	44.51	4634.79	-2318.85
Nonignorable						
N(0, 1)	109.13	4602.19	-2322.10	117.11	4539.79	-2313.84
N(0, 2)	133.85	4575.71	-2321.11	149.19	4451.91	-2310.67
N(0, 3)	143.11	4543.59	-2319.55	169.03	4396.48	-2309.39
N(0, 4)	157.89	4544.77	-2319.87	166.84	4315.97	-2306.42
N(0, 5)	167.68	4536.66	-2319.71	174.49	4288.92	-2305.63
N(0, 6)	167.47	4511.15	-2318.84	168.71	4265.32	-2304.53
N(0, 7)	179.29	4522.07	-2319.48	186.35	4248.50	-2302.97
N(0, 8)	177.36	4505.24	**-2317.75**	195.31	4248.02	-2304.15
N(0, 9)	187.83	**4496.09**	-2318.87	185.49	4224.16	**-2301.43**
N(0, 10)	187.02	4498.76	-2318.64	184.69	**4187.97**	-2302.40
Jeffreys Prior	180.86	4550.19	**-2318.56**	176.33	4312.23	**-2302.26**

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 of the count data are shown on the first three columns in Table 4.6. The effective number of parameters under the ignorable model ($p_D = 44.58$) 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}. We also see from Table 4.6 that (i) the DIC value was 4634.93 under the ignorable model; (ii) under the nonignorable model with a $N(0, \sigma^2_{prior})$ prior, the value of DIC first decreased and slightly increased as σ^2_{prior} increased; (iii) the DIC attained the local minimum with DIC=4496.09 at $\sigma^2_{prior} = 9$ among all the models under consideration (10 values of σ^2_{prior} and Jeffreys Prior).

The results indicated by LPML were slightly different from the results by the DIC criterion. The nonignorable model with a $N(0, 8)$ prior had the largest value of LPML (LPML=-2317.75) among all the models under consideration. The nonignorable model with Jeffreys prior had the second largest value of LPML (LPML=-2318.56). These results indicate that for the HIV prevention behavioral data, the missing longitudinal binary responses were potentially nonignorably missing.

We further see from Table 4.6 that (i) under the ignorable model, the DIC (4634.93) and LPML (-2318.96) values of the binary response variable were similar to the DIC (4634.79) and LPML (-2318.85) values of the Poisson response variable; (ii) under the nonignorable models, all of the DIC values of the Poisson response variable were smaller than those of the binary response variable, indicating that the missing data mechanism of the Poisson response variable performed better; (iii) under the nonignorable models, all of the LPML values of the Poisson response variable were larger than those of the binary response variable, again, implying that the missing data mechanism of the Poisson response variable performed better. These findings confirmed our conjecture that the Poisson response variable contained more information than the binary response variable and therefore should improve the fit of the missing data mechanism.

Table 4.7: Posterior Summaries under the Ignorable Model for the HIV Prevention Behavioral Data.

	Binary Response Model		
	EST	SD	95% HPD Interval
Baseline			
Intercept	-0.699	0.200	(-1.102, -0.322)
Gender	0.384	0.132	(0.129, 0.644)
City	0.124	0.157	(-0.190, 0.427)
Cohabit	0.724	0.142	(0.458, 1.008)
Counselor	0.433	0.161	(0.121, 0.753)
Drink	0.435	0.352	(-0.253, 1.127)
Age	-0.372	0.073	(-0.510, -0.227)
6-Month			
Intercept	-1.776	0.268	(-2.301, -1.273)
Gender	0.158	0.141	(-0.122, 0.431)
City	0.115	0.168	(-0.228, 0.433)
Cohabit	0.643	0.145	(0.367, 0.930)
Counselor	0.581	0.181	(0.216, 0.927)
Drink	0.999	0.370	(0.283, 1.733)
Age	-0.463	0.082	(-0.626, -0.307)
12-Month			
Intercept	-1.824	0.280	(-2.385, -1.293)
Gender	0.334	0.151	(0.035, 0.628)
City	-0.006	0.177	(-0.369, 0.332)
Cohabit	0.646	0.151	(0.352, 0.940)
Counselor	0.280	0.183	(-0.085, 0.633)
Drink	0.593	0.367	(-0.113, 1.325)
Age	-0.491	0.088	(-0.665, -0.322)
18-Month			
Intercept	-1.759	0.279	(-2.315, -1.233)
Gender	0.246	0.149	(-0.053, 0.533)
City	-0.147	0.180	(-0.509, 0.198)
Cohabit	0.495	0.148	(0.206, 0.782)
Counselor	0.412	0.188	(0.041, 0.783)
Drink	0.585	0.381	(-0.155, 1.340)
Age	-0.399	0.083	(-0.560, -0.235)
z			
Baseline	0.087	0.121	(-0.151, 0.323)
6-Month	-0.156	0.130	(-0.414, 0.098)
12-Month	-0.430	0.142	(-0.714, -0.156)
18-Month	-0.374	0.143	(-0.662, -0.105)
ρ	0.790	0.035	(0.721, 0.858)
σ^2	3.037	0.814	(1.747, 4.603)
τ	1.068	1.237	(0.000, 3.638)

	Missing Data Mechanism (Intermittent Missing)		
	EST	SD	95% HPD Interval
Baseline			
Intercept	-3.452	0.419	(-4.290, -2.648)
Gender	0.097	0.237	(-0.352, 0.574)
City	-0.307	0.328	(-0.942, 0.336)
Cohabit	0.248	0.228	(-0.198, 0.695)
Counselor	0.679	0.373	(-0.046, 1.406)
Age	0.073	0.112	(-0.142, 0.296)
6-Month			
Intercept	-2.545	0.302	(-3.147, -1.973)
Gender	-0.461	0.193	(-0.825, -0.073)
City	-0.342	0.266	(-0.882, 0.165)
Cohabit	0.022	0.187	(-0.343, 0.392)
Counselor	0.368	0.266	(-0.140, 0.898)
Age	-0.029	0.096	(-0.222, 0.151)
R_0	0.928	0.308	(0.326, 1.528)
12-Month			
Intercept	-2.402	0.307	(-3.023, -1.823)
Gender	-0.519	0.220	(-0.943, -0.086)
City	-0.669	0.340	(-1.354, -0.022)
Cohabit	-0.160	0.214	(-0.560, 0.280)
Counselor	-0.515	0.243	(-0.997, -0.045)
Age	-0.065	0.111	(-0.282, 0.151)
R_0	0.252	0.466	(-0.684, 1.127)
R_1	1.023	0.300	(0.423, 1.601)
z			
Baseline	-0.630	0.231	(-1.090, -0.189)
6-Month	-0.111	0.184	(-0.463, 0.257)
12-Month	0.665	0.224	(0.219, 1.084)

	Missing Data Mechanism (Dropout)		
	EST	SD	95% HPD Interval
6-Month			
Intercept	-3.209	0.327	(-3.865, -2.579)
Gender	-0.336	0.217	(-0.759, 0.091)

	Missing Data Mechanism (Dropout)		
	EST	SD	95% HPD Interval
12-Month			
Intercept	-3.049	0.376	(-3.808, -2.340)
Gender	-0.670	0.257	(-1.182, -0.172)

City	0.425	0.245	(-0.076, 0.884)	City	-0.244	0.341	(-0.925, 0.411)
Cohabit	0.432	0.215	(0.011, 0.855)	Cohabit	-0.353	0.252	(-0.838, 0.154)
Counselor	0.147	0.276	(-0.373, 0.701)	Counselor	0.234	0.332	(-0.395, 0.898)
Age	-0.231	0.116	(-0.464, -0.011)	Age	-0.208	0.132	(-0.466, 0.054)
18-Month				**z**			
Intercept	-2.555	0.286	(-3.128, -2.001)	6-Month	0.013	0.209	(-0.411, 0.408)
Gender	-0.474	0.195	(-0.847, -0.087)	12-Month	0.281	0.248	(-0.205, 0.766)
City	0.352	0.225	(-0.100, 0.775)	18-Month	0.053	0.186	(-0.312, 0.413)
Cohabit	0.191	0.190	(-0.174, 0.567)				
Counselor	0.082	0.245	(-0.393, 0.566)				
Age	-0.063	0.099	(-0.262, 0.123)				

Table 4.8: Posterior Summaries under the Nonignorable Model with a $N(0, 8)$ Prior for the HIV Prevention Behavioral Data.

	Poisson Response Model				Missing Data Mechanism (Intermittent Missing)		
	EST	SD	95% HPD Interval		EST	SD	95% HPD Interval
Baseline				**Baseline**			
Intercept	-0.653	0.189	(-1.014, -0.276)	Intercept	-3.705	0.672	(-5.006, -2.481)
Gender	0.364	0.125	(0.129, 0.617)	Gender	0.084	0.239	(-0.370, 0.571)
City	0.116	0.149	(-0.169, 0.413)	City	-0.302	0.326	(-0.937, 0.338)
Cohabit	0.694	0.133	(0.443, 0.964)	Cohabit	0.212	0.249	(-0.263, 0.711)
Counselor	0.422	0.153	(0.123, 0.721)	Counselor	0.648	0.374	(-0.072, 1.384)
Drink	0.416	0.336	(-0.242, 1.076)	Age	0.093	0.124	(-0.143, 0.344)
Age	-0.353	0.068	(-0.490, -0.227)	y_0			
6-Month				**6-Month**			
Intercept	-1.645	0.265	(-2.157, -1.125)	Intercept	-2.627	0.342	(-3.280, -1.952)
Gender	0.130	0.134	(-0.125, 0.400)	Gender	-0.458	0.193	(-0.836, -0.086)
City	0.091	0.160	(-0.226, 0.403)	City	-0.340	0.271	(-0.874, 0.178)
Cohabit	0.598	0.139	(0.332, 0.878)	Cohabit	-0.003	0.195	(-0.371, 0.389)
Counselor	0.545	0.169	(0.221, 0.885)	Counselor	0.339	0.272	(-0.183, 0.887)
Drink	0.934	0.344	(0.264, 1.609)	Age	-0.003	0.106	(-0.206, 0.209)
Age	-0.433	0.078	(-0.598, -0.291)	R_0	0.928	0.318	(0.303, 1.547)
				y_0	-0.159	0.320	(-0.793, 0.455)
12-Month				y_1	0.374	0.769	(-1.031, 1.976)
Intercept	-1.671	0.273	(-2.209, -1.144)				
Gender	0.317	0.147	(0.031, 0.608)	**12-Month**			
City	-0.027	0.167	(-0.353, 0.302)	Intercept	-2.646	0.367	(-3.379, -1.944)
Cohabit	0.600	0.142	(0.315, 0.868)	Gender	-0.580	0.231	(-1.047, -0.141)
Counselor	0.215	0.171	(-0.133, 0.541)	City	-0.680	0.338	(-1.357, -0.042)
Drink	0.527	0.344	(-0.146, 1.200)	Cohabit	-0.217	0.222	(-0.632, 0.236)
Age	-0.452	0.082	(-0.613, -0.292)	Counselor	-0.533	0.247	(-1.022, -0.056)
				Age	-0.021	0.117	(-0.244, 0.209)
18-Month				R_0	0.320	0.480	(-0.637, 1.241)
Intercept	-1.414	0.262	(-1.946, -0.934)	R_1	1.040	0.311	(0.445, 1.652)
Gender	0.142	0.141	(-0.130, 0.423)	y_1	-0.268	0.389	(-1.028, 0.497)
City	-0.060	0.172	(-0.403, 0.269)	y_2	0.908	0.706	(-0.461, 2.268)
Cohabit	0.453	0.136	(0.191, 0.724)				
Counselor	0.351	0.172	(0.028, 0.698)	z			
Drink	0.544	0.352	(-0.139, 1.234)	Baseline	-0.640	0.235	(-1.112, -0.194)
Age	-0.353	0.078	(-0.509, -0.206)	6-Month	-0.093	0.188	(-0.460, 0.278)
				12-Month	0.750	0.234	(0.298, 1.217)

z			
Baseline	0.075	0.116	(-0.153, 0.301)
6-Month	-0.147	0.123	(-0.392, 0.091)
12-Month	-0.395	0.136	(-0.665, -0.131)
18-Month	-0.319	0.131	(-0.567, -0.054)
ρ	0.792	0.037	(0.721, 0.862)
σ^2	2.653	0.692	(1.543, 4.003)
τ	1.121	1.275	(0.000, 3.770)

	Missing Data Mechanism (Dropout)				Missing Data Mechanism (Dropout)		
	EST	SD	95% HPD Interval		EST	SD	95% HPD Interval
6-Month				**12-Month**			
Intercept	-3.214	0.428	(-4.073, -2.424)	Intercept	-2.931	0.420	(-3.788, -2.161)
Gender	-0.334	0.222	(-0.778, 0.101)	Gender	-0.631	0.263	(-1.141, -0.115)
City	0.450	0.251	(-0.027, 0.952)	City	-0.238	0.342	(-0.917, 0.417)
Cohabit	0.469	0.230	(0.006, 0.907)	Cohabit	-0.278	0.262	(-0.788, 0.244)
Counselor	0.176	0.294	(-0.419, 0.731)	Counselor	0.254	0.332	(-0.368, 0.931)
Age	-0.254	0.140	(-0.522, 0.032)	Age	-0.253	0.140	(-0.543, 0.009)
y_0	-0.112	0.443	(-1.034, 0.637)	y_1	0.024	0.461	(-0.963, 0.840)
y_1	-0.837	2.022	(-4.998, 2.868)	y_2	-2.043	2.033	(-6.386, 1.603)
18-Month				**z**			
Intercept	-3.784	1.058	(-6.038, -2.241)	6-Month	0.001	0.222	(-0.438, 0.428)
Gender	-0.590	0.216	(-1.009, -0.166)	12-Month	0.211	0.257	(-0.284, 0.725)
City	0.446	0.246	(-0.026, 0.932)	18-Month	0.127	0.204	(-0.268, 0.531)
Cohabit	0.075	0.208	(-0.319, 0.493)				
Counselor	-0.124	0.273	(-0.655, 0.415)				
Age	0.040	0.110	(-0.173, 0.259)				
y_2	-0.916	0.377	(-1.606, -0.175)				
y_3	2.910	1.445	(0.437, 6.101)				

Table 4.9: Posterior Summaries under the Nonignorable Model with Jeffreys Prior for the HIV Prevention Behavioral Data.

	Poisson Response Model				Missing Data Mechanism (Intermittent Missing)		
	EST	SD	95% HPD Interval		EST	SD	95% HPD Interval
Baseline				**Baseline**			
Intercept	-0.664	0.188	(-1.035, -0.304)	Intercept	-3.520	0.575	(-4.651, -2.480)
Gender	0.368	0.127	(0.112, 0.614)	Gender	0.088	0.238	(-0.370, 0.557)
City	0.116	0.152	(-0.175, 0.417)	City	-0.268	0.315	(-0.886, 0.338)
Cohabit	0.697	0.133	(0.438, 0.956)	Cohabit	0.226	0.241	(-0.252, 0.692)
Counselor	0.422	0.153	(0.120, 0.725)	Counselor	0.610	0.368	(-0.107, 1.342)
Drink	0.424	0.338	(-0.231, 1.086)	Age	0.085	0.122	(-0.161, 0.315)
Age	-0.356	0.069	(-0.491, -0.222)	y_0	0.211	0.770	(-1.176, 1.806)
6-Month				**6-Month**			
Intercept	-1.671	0.266	(-2.184, -1.160)	Intercept	-2.558	0.332	(-3.216, -1.930)
Gender	0.134	0.137	(-0.134, 0.403)	Gender	-0.453	0.191	(-0.827, -0.083)
City	0.097	0.164	(-0.227, 0.416)	City	-0.319	0.264	(-0.847, 0.187)
Cohabit	0.605	0.141	(0.335, 0.885)	Cohabit	0.007	0.195	(-0.366, 0.393)

Counselor	0.551	0.170	(0.216, 0.882)
Drink	0.952	0.351	(0.258, 1.638)
Age	-0.438	0.079	(-0.592, -0.287)
12-Month			
Intercept	-1.685	0.277	(-2.215, -1.150)
Gender	0.320	0.148	(0.023, 0.600)
City	-0.027	0.171	(-0.361, 0.310)
Cohabit	0.601	0.145	(0.320, 0.883)
Counselor	0.220	0.173	(-0.115, 0.564)
Drink	0.541	0.347	(-0.153, 1.204)
Age	-0.457	0.084	(-0.629, -0.304)
18-Month			
Intercept	-1.463	0.276	(-2.021, -0.954)
Gender	0.154	0.144	(-0.124, 0.442)
City	-0.073	0.176	(-0.420, 0.270)
Cohabit	0.456	0.140	(0.178, 0.721)
Counselor	0.365	0.174	(0.031, 0.708)
Drink	0.555	0.351	(-0.119, 1.273)
Age	-0.362	0.080	(-0.520, -0.212)
z			
Baseline	0.080	0.118	(-0.154, 0.311)
6-Month	-0.148	0.125	(-0.401, 0.092)
12-Month	-0.400	0.138	(-0.672, -0.133)
18-Month	-0.328	0.135	(-0.588, -0.056)
ρ	0.791	0.037	(0.721, 0.866)
σ^2	2.729	0.751	(1.601, 4.186)
τ	1.106	1.281	(0.000, 3.846)

Counselor	0.320	0.271	(-0.205, 0.861)
Age	-0.006	0.105	(-0.211, 0.196)
R_0	0.901	0.314	(0.260, 1.494)
y_0	-0.125	0.303	(-0.730, 0.448)
y_1	0.279	0.713	(-1.071, 1.694)
12-Month			
Intercept	-2.581	0.362	(-3.286, -1.873)
Gender	-0.573	0.228	(-1.008, -0.109)
City	-0.635	0.328	(-1.294, -0.012)
Cohabit	-0.212	0.221	(-0.653, 0.208)
Counselor	-0.542	0.246	(-1.026, -0.057)
Age	-0.021	0.116	(-0.244, 0.210)
R_0	0.387	0.467	(-0.529, 1.281)
R_1	0.976	0.306	(0.370, 1.573)
y_1	-0.244	0.384	(-0.993, 0.508)
y_2	0.870	0.685	(-0.480, 2.207)
z			
Baseline	-0.629	0.230	(-1.076, -0.184)
6-Month	-0.095	0.187	(-0.464, 0.268)
12-Month	0.735	0.232	(0.279, 1.185)

	Missing Data Mechanism (Dropout)		
	EST	SD	95% HPD Interval
6-Month			
Intercept	-3.141	0.400	(-3.932, -2.388)
Gender	-0.326	0.220	(-0.760, 0.100)
City	0.458	0.243	(0.000, 0.949)
Cohabit	0.458	0.228	(0.016, 0.908)
Counselor	0.149	0.289	(-0.407, 0.725)
Age	-0.246	0.138	(-0.519, 0.020)
y_0	-0.109	0.422	(-0.938, 0.652)
y_1	-0.750	1.944	(-4.945, 2.634)
18-Month			
Intercept	-3.509	1.043	(-5.478, -2.040)
Gender	-0.566	0.213	(-0.984, -0.151)
City	0.445	0.241	(-0.028, 0.919)
Cohabit	0.085	0.205	(-0.315, 0.486)
Counselor	-0.116	0.264	(-0.630, 0.408)
Age	0.031	0.111	(-0.187, 0.248)
y_2	-0.798	0.444	(-1.614, 0.115)
y_3	2.466	1.605	(-0.516, 5.926)

	Missing Data Mechanism (Dropout)		
	EST	SD	95% HPD Interval
12-Month			
Intercept	-2.872	0.409	(-3.686, -2.077)
Gender	-0.615	0.260	(-1.131, -0.116)
City	-0.202	0.342	(-0.896, 0.441)
Cohabit	-0.278	0.262	(-0.798, 0.226)
Counselor	0.223	0.334	(-0.447, 0.860)
Age	-0.240	0.139	(-0.516, 0.031)
y_1	0.021	0.460	(-0.959, 0.835)
y_2	-1.874	2.108	(-6.306, 1.928)
z			
6-Month	0.002	0.218	(-0.423, 0.429)
12-Month	0.209	0.256	(-0.278, 0.724)
18-Month	0.119	0.202	(-0.272, 0.512)

Chapter 5

Extension and Future Research Direction

5.1 Extensions to Choice of Link Functions

In Section 2.1.1, we assume the probit link function for the binary response variable, the logit link function for the missing data mechanisms in Section 2.1.2 and 3.1.2, and the logit and log links for modeling the count response variable in Section 3.1.1. However, other choices of link can also be applied as long as the transformation is one-to-one, continuous and differentiable. For example, we may use symmetric link function such as the robit link (Liu, 2004). It is well known that the robit regression model provides a rich class of models, including logit and probit regression models as special cases, for analysis of binary response data. However, the symmetric link has an inferior performance when the data structure requires a skewed response probability function. In this case, we may use the asymmetric link such as the complementary log-log link when the response probability function is positively skewed. Other link functions such as Stukel's models (Stukel, 1988) and generalized skewed t-link models Kim *et al.* (2008) are more general and can be applied to any types of skewed distribution for the binary response data.

Similarly, for the count response data, instead of using the zero-inflated Poisson model, we can use the zero-inflated negative binomial model.

5.2 Extensions to Missing Data Mechanism and Multiple Longitudinal Outcomes

5.2.1 Path Specific Model

Due to the repeated measure property of longitudinal study, it is very common to encounter numerous missing visits. The person may permanently withdraw from the study or may just miss one or several scheduled visits, but will finally return to the study. The missingness can be thus categorized into two types, i.e., dropout and intermittent missing, which can be uniquely determined by the person's entire visiting path. However, we never know which type the missingness belongs to until we witness the entire path. Given this built-in nature and constraint, a retrospective model, which models the past event based on the current or future information, is easier to implement. However, it is more desirable if we can model the missing status forward, especially for the purpose of prediction.

For a length of T visit clinical trial, the total number of potential paths is 2^T and 2^{T-1} if we assume no missingness at baseline. Since the value of T is usually small due to the financial or other reasons, it is entirely possible to model the missingness of each path. We thus propose the idea of the path specific prospective model.

Recall that $\mathbf{R}_T = (R_1, \dots, R_T)'$ denotes the vector of the missing data indicators as well as the entire missing path, where R_t is defined as

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

Assume all the baseline measurements are observed ($\mathbf{R}_0 = 0$). We propose the following model,

$$P(R_1 = r_1, \dots, R_T = r_T | z, \mathbf{X}, \mathbf{y}_T, \boldsymbol{\gamma}) = \frac{\psi(r_1, \dots, r_T | z, \mathbf{X}, \mathbf{y}_T, \boldsymbol{\gamma})}{\sum_{\mathbf{R}'_T \in \Omega} \psi(\mathbf{R}'_T | z, \mathbf{X}, \mathbf{y}_T, \boldsymbol{\gamma})}, \tag{5.1}$$

where Ω is the set of the exhaustive paths, and the path specific function ψ is given as follows,

$$\psi(\mathbf{R}_T | z, \mathbf{X}, \mathbf{y}_T, \boldsymbol{\gamma}) = \exp(z\gamma_{1\mathbf{R}_T} + \mathbf{x}'\boldsymbol{\gamma}_{2\mathbf{R}_T} + g(\mathbf{y}_T, \boldsymbol{\gamma}_{3\mathbf{R}_T})), \tag{5.2}$$

where $\mathbf{x}$ is the baseline covariates, g is a certain linear function of $\mathbf{y}_t$ and the corresponding coefficients $\boldsymbol{\gamma}_{3\mathbf{R}_T}$. In addition, all the coefficients $\boldsymbol{\gamma}_{\mathbf{R}_T} = (\gamma_{1\mathbf{R}_T}, \boldsymbol{\gamma}_{2\mathbf{R}_T}, \boldsymbol{\gamma}_{3\mathbf{R}_T})$ are path specific. Moreover, assume $\tilde{\mathbf{R}}_T$ is the path which has the fewest observations. We set $\tilde{\mathbf{R}}_T$ as our baseline path, and let $\psi(\tilde{\mathbf{R}}_T | z, \mathbf{X}, \mathbf{y}_T, \boldsymbol{\gamma}) = 1$.

Remark 5.1: We can easily obtain the marginal probability or joint probability of any combination of $(R_1, \dots, R_T)$ by summing over all the other irrelevant missing data indicators. For better illustration,

$$P(R_t = r_t) = \sum_{\mathbf{R}'_T \in \Omega^*} P(\mathbf{R}'_T),$$

where $\Omega^* = \{\mathbf{R}'_T |$ the t_{th} missing indicator is $r_t, t > 0\}$.

Remark 5.2: If we assume that $P(R_T | z, \mathbf{X}, \mathbf{y}_T, \boldsymbol{\gamma})$ depends on the longitudinal measures through visit t, we simply take $h(\mathbf{y}_t, \boldsymbol{\gamma}_{3\mathbf{R}_T}) = \boldsymbol{\gamma}'_{3\mathbf{R}_t}\mathbf{y}_t$ in (5.2). The model in (5.1) is thus nonignorable due to the existence of intermittent missing and dropouts. We can also let $h(\mathbf{y}_t, \boldsymbol{\gamma}_{3\mathbf{R}_T}) = 0$ if we assume ignorable missing.

Remark 5.3: It may not be necessary to specify 2^{T-1} different models for each path. The model for some path may coincide with the model for the other path. In other word, $\boldsymbol{\gamma}_{R_T}$ is equal to $\boldsymbol{\gamma}_{R'_T}$, for $R_T \neq R'_T$. The equivalence of the coefficients for different paths can be easily tested via various methods, such as Bayes factor from the Bayesian perspective.

5.2.2 Joint Modeling of Multiple Types of Responses

In the HIV prevention behavioral intervention clinical trial, several different types of responses were collected repeatedly over time on the same subjects. According to Fisher *et al.* (2014), the primary outcome measures for intervention evaluation were ACASI-reported number of sexual events without condoms over the past 4 weeks with all partners, regardless of perceived partner serostatus, and number of sexual events without condoms over the past 4 weeks with partners perceived to be HIV negative or HIV-status unknown. Additional outcome measures included interviewer collected information on number of sexual events without condom use (past 4 weeks), a Likert item on consistency of condom use (during the past 4 weeks, and past 3 months), and whether or not last condom nonuse was in the past 6 months. These partially overlapping interviewer-delivered measures were included to provide multiple, potentially convergent end points assessed through alternative methodologies (ACASI and interviewer) over varying time periods.

Our future research direction includes developing a joint model for different types of response measurements as well as the corresponding missing data mechanisms. The joint model is more desirable since it allows for borrowing strength among these longitudinal responses. One possible approach to capture the dependence between different types of responses is by introducing the random effects.

Another possible way to joint modeling different types of measurements is by using the technique of copula modeling. Copula, which is defined as the joint cumulative distribution function of multiple random variables, allows us to capture the dependence of different random variables, and obtain the joint distribution by estimating marginal distributions and copula separately. Since most of our response measurements are discrete, copulas for discrete data are more suitable, which, however, are far more challenging than the copulas for the continuous data. As elaborated in Genest and Nešlehová (2007), copula modeling is feasible for constructing multivariate distributions with discrete margins, but modeling and interpreting dependence through copulas is subject to caution. First of all, the uniqueness property of the copula no longer exists. Consequently, the copula alone does not characterize the dependence between two random variables. Also, conventional measures of concordance such as Kendall's τ or Spearman's ρ, which provide margin-free measures of the level of dependence in the continuous bivariate distribution, do not maintain the nice property when random variables are discrete. In conclusion, copula modeling for discrete data is a challenging but desirable approach.

5.3 Extensions to Model Selection Criteria

We currently use the DIC and conditional LPML criteria to assess fit of the missing data mechanism. 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.*, 2017). The decomposition of DIC and LPML are more challenging in this setting due to the involvement of discrete data. More effective computational algorithms need to be developed, which requires further exploration.

Appendix A

Proofs of Theorems

Proof of Proposition 2.2.1. 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

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

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 2.2.2. 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

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

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

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

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 $\boldsymbol{\epsilon}_i^{*\prime}\Sigma\boldsymbol{\epsilon}_i^* \leq \lambda_{max}\|\boldsymbol{\epsilon}_i^*\|^2$. Therefore,

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

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 $(A.1)$.

Proof of Proposition 3.2.3. If we assume $\pi(\boldsymbol{\beta}_1, \boldsymbol{\gamma}_1, \boldsymbol{\alpha}, \boldsymbol{\phi}) = 1$

$$\begin{aligned}
&\pi^*(\boldsymbol{\theta}|D_{\text{obs}}) = \mathcal{L}(\boldsymbol{\theta}|D_{\text{obs}})\pi(\boldsymbol{\beta}_2, \boldsymbol{\gamma}_2, \tau, \sigma^2, \rho) \\
=&\sum_{y_{\text{mis}}} \prod_{i=1}^{n} \prod_{k=1}^{K} \Big\{ \int f_y(\mathbf{y}_i|z_i, \mathbf{x}_i, k_i, \zeta_{k_i}, \boldsymbol{\epsilon}_i, \boldsymbol{\theta}) f(\boldsymbol{\epsilon}_i|\alpha, \rho) d\boldsymbol{\epsilon} f(\zeta_k|\tau) d\boldsymbol{\zeta} \\
&f_{W|\mathbf{y}}(w_i|z_i, \mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\alpha}) f_{\mathbf{R}|W,\mathbf{y}}(\mathbf{R}_{iw_i-2}|z_i, \mathbf{x}_i, \mathbf{y}_i, w_i, \boldsymbol{\phi}) \\
&\pi(\boldsymbol{\beta}_2, \boldsymbol{\gamma}_2, \tau, \sigma^2, \rho) \Big\} \\
\geq&\sum_{y_{\text{mis}}} \prod_{i=1}^{n} \prod_{k=1}^{K} \Big\{ \int \prod_{t=0}^{T} (1 - \pi_{it}) \frac{\mu_{it}^{y_{it}} e^{-\mu_{it}}}{y_{it}!} f(\boldsymbol{\epsilon}_i|\alpha, \rho) d\boldsymbol{\epsilon} f(\zeta_k|\tau) d\boldsymbol{\zeta} \\
&f_{W|\mathbf{y}}(w_i|z_i, \mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\alpha}) f_{\mathbf{R}|W,\mathbf{y}}(\mathbf{R}_{iw_i-2}|z_i, \mathbf{x}_i, \mathbf{y}_i, w_i, \boldsymbol{\phi}) \\
&\pi(\boldsymbol{\beta}_2, \boldsymbol{\gamma}_2, \tau, \sigma^2, \rho) \Big\}
\end{aligned}$$

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

$$\begin{aligned}
\pi^*(\boldsymbol{\theta}|D_{\text{obs}}) \geq &\prod_{i=1}^{n} \prod_{k=1}^{K} \Big\{ \int \prod_{t=0}^{T} (1 - \pi_{it}) \frac{\mu_{it}^{y_{it}^*} e^{-\mu_{it}}}{y_{it}^*!} f(\boldsymbol{\epsilon}_i|\alpha, \rho) d\boldsymbol{\epsilon} f(\zeta_k|\tau) d\boldsymbol{\zeta} \\
&f_{W|\mathbf{y}}(W_i = 1|z_i, \mathbf{x}_i, \mathbf{y}_i^*, \boldsymbol{\alpha}_1)^{\mathbf{1}(w_i=1)} \\
&\Big\{(1 - f_{W|\mathbf{y}}(W_i = 1|z_i, \mathbf{x}_i, \mathbf{y}_i^*, \boldsymbol{\alpha}_1)) \\
&\prod_{\ell=2}^{w_i-1} (1 - f_{W|\mathbf{y}}(W_i = \ell|W_i > \ell - 1, z_i, \mathbf{x}_i, \mathbf{y}_i^*, \boldsymbol{\alpha}_\ell)) \Big\}^{\mathbf{1}(w_i>1)} \\
&f_{W|\mathbf{y}}(W_i = w_i|W_i > w_i - 1, z_i, \mathbf{x}_i, \mathbf{y}_i^*, \boldsymbol{\alpha}_{w_i})^{\mathbf{1}(1<w_i<T+1)} \\
&\prod_{j=0}^{w_i-2} f_{\mathbf{R}|W,\mathbf{y}}(R_{ij}|\mathbf{R}_{ij-1}, z_i, \mathbf{x}_i, \mathbf{y}_i^*, w_i, \boldsymbol{\phi}_j) \pi(\boldsymbol{\beta}_2, \boldsymbol{\gamma}_2, \tau, \sigma^2, \rho) \Big\}.
\end{aligned}$$

Since the unnormalized marginal posterior density of $(\boldsymbol{\beta}_1, \boldsymbol{\gamma}_1)$ with $\pi(\boldsymbol{\beta}_1, \boldsymbol{\gamma}_1) = 1$ is $(1 - \pi_{it})$, where π_{it} is given in (3.1), the joint posterior is obviously improper. Note that for each j, the unnormalized marginal posterior density of $\boldsymbol{\phi}_j$ with $\pi(\boldsymbol{\phi}_j) = 1$ is

$\prod_{i=1}^{n} f_{\mathbf{R}|W,\mathbf{y}}(R_{ij}|\mathbf{R}_{ij-1}, z_i, \mathbf{x}_i, \mathbf{y}_i^*, w_i, \boldsymbol{\phi}_j)^{\mathbf{1}(j \leq w_i - 2)}$, which corresponds to a binary regression model with response equal to R_{ij}. Due to the construction of y_i^* and Proposition A.1 (Huang *et al.*, 2005), the posterior density of $\boldsymbol{\phi}_j$ is improper and thus the joint posterior $\pi^*(\boldsymbol{\theta}|D_{\text{obs}})$ is also improper. Similarly, the joint posterior $\pi^*(\boldsymbol{\theta}|D_{\text{obs}})$ is improper if we take $\pi(\boldsymbol{\alpha}_\ell) = 1$.

Proof of Proposition 3.2.4. Following the thread of proof in Proposition 2.2.2 and Theorem 3.1 in Chen *et al.* (2002), we can easily obtain the result.

Appendix B

Additional Tables

Table B.1: Posterior Summaries under the Nonignorable Model with a N(0, 1) Prior and a N(0, 10) Prior when missing percentage is low.

		N(0, 1) Prior					N(0, 10) Prior				
	TRUE	EST	SE	SIME	RMSE	CP	EST	SE	SIME	RMSE	CP
t=0											
β^*_{00}	-1.000	-1.031	0.133	0.121	0.124	0.984	-1.003	0.136	0.123	0.123	0.968
β^*_{10}	0.500	0.512	0.069	0.069	0.070	0.968	0.504	0.068	0.068	0.068	0.956
β^*_{20}	1.000	1.022	0.134	0.129	0.131	0.968	0.999	0.132	0.128	0.127	0.956
β^*_{30}	0.400	0.410	0.111	0.100	0.100	0.980	0.400	0.110	0.098	0.098	0.984
γ_{00}	-2.500	-2.573	0.239	0.236	0.247	0.964	-2.717	0.441	0.442	0.492	0.960
γ_{10}	0.500	0.509	0.118	0.108	0.108	0.972	0.499	0.128	0.119	0.118	0.964
γ_{20}	-0.500	-0.469	0.235	0.222	0.224	0.964	-0.490	0.254	0.246	0.245	0.956
γ_{30}	-0.500	-0.493	0.213	0.203	0.202	0.964	-0.503	0.218	0.205	0.204	0.972
γ_{60}	0.000	-0.084	0.581	0.468	0.474	0.992	0.040	0.972	0.843	0.843	0.976
t=1											
β^*_{01}	-1.000	-1.048	0.155	0.155	0.162	0.964	-0.987	0.168	0.176	0.176	0.912
β^*_{11}	0.500	0.499	0.073	0.068	0.068	0.964	0.498	0.072	0.068	0.068	0.972
β^*_{21}	1.000	1.017	0.141	0.140	0.140	0.964	0.977	0.143	0.142	0.143	0.928
β^*_{31}	-0.200	-0.193	0.111	0.105	0.105	0.956	-0.196	0.110	0.103	0.103	0.952
γ_{01}	-2.000	-2.012	0.210	0.174	0.175	0.972	-2.218	0.392	0.454	0.503	0.956
γ_{11}	0.500	0.521	0.088	0.091	0.093	0.932	0.503	0.095	0.098	0.098	0.928
γ_{21}	-0.500	-0.475	0.179	0.176	0.178	0.968	-0.519	0.194	0.200	0.201	0.952
γ_{31}	-0.250	-0.279	0.159	0.151	0.154	0.956	-0.259	0.164	0.155	0.155	0.964
γ_{41}	0.400	0.376	0.285	0.293	0.293	0.944	0.394	0.298	0.303	0.303	0.948
γ_{51}	-0.250	-0.169	0.263	0.234	0.247	0.948	-0.268	0.305	0.309	0.308	0.936
γ_{61}	0.500	0.266	0.625	0.514	0.564	0.968	0.599	0.960	1.014	1.017	0.944
t=2											
β^*_{02}	-1.000	-1.067	0.149	0.148	0.162	0.964	-1.009	0.154	0.156	0.156	0.960
β^*_{12}	0.500	0.495	0.072	0.067	0.068	0.968	0.497	0.071	0.067	0.067	0.972
β^*_{22}	1.000	1.041	0.145	0.137	0.143	0.964	1.000	0.145	0.138	0.138	0.964
β^*_{32}	-0.400	-0.405	0.116	0.111	0.111	0.948	-0.394	0.114	0.109	0.109	0.952

	TRUE	EST	SE	SIME	RMSE	CP	EST	SE	SIME	RMSE	CP
γ_{02}	-2.800	-2.783	0.238	0.243	0.243	0.948	-2.970	0.338	0.414	0.446	0.932
γ_{12}	0.500	0.515	0.087	0.093	0.094	0.940	0.500	0.090	0.098	0.098	0.940
γ_{22}	-0.500	-0.475	0.179	0.161	0.163	0.956	-0.529	0.190	0.180	0.182	0.964
γ_{32}	0.250	0.244	0.161	0.170	0.170	0.932	0.269	0.165	0.178	0.179	0.928
γ_{42}	1.700	1.704	0.163	0.168	0.168	0.944	1.764	0.181	0.197	0.207	0.944
γ_{52}	-0.600	-0.525	0.257	0.275	0.284	0.916	-0.624	0.274	0.308	0.308	0.896
γ_{62}	1.300	1.070	0.492	0.473	0.525	0.928	1.417	0.638	0.740	0.748	0.940
t=3											
β^*_{03}	-1.000	-1.029	0.140	0.136	0.138	0.944	-1.001	0.143	0.140	0.140	0.944
β^*_{13}	0.500	0.506	0.077	0.080	0.080	0.940	0.501	0.076	0.080	0.080	0.944
β^*_{23}	1.000	1.027	0.142	0.134	0.137	0.956	1.003	0.140	0.131	0.131	0.960
β^*_{33}	-0.600	-0.614	0.124	0.122	0.122	0.944	-0.602	0.122	0.120	0.120	0.948
γ_{03}	-2.800	-2.872	0.178	0.187	0.200	0.924	-2.896	0.192	0.203	0.224	0.916
γ_{13}	0.500	0.501	0.089	0.093	0.092	0.940	0.499	0.092	0.098	0.098	0.936
γ_{23}	-0.500	-0.494	0.171	0.165	0.165	0.956	-0.500	0.176	0.172	0.171	0.972
γ_{33}	0.500	0.515	0.162	0.168	0.169	0.936	0.519	0.165	0.174	0.174	0.936
γ_{43}	1.700	1.752	0.115	0.118	0.129	0.944	1.750	0.119	0.123	0.132	0.940
γ_{53}	0.600	0.591	0.247	0.227	0.226	0.964	0.578	0.264	0.263	0.264	0.940
γ_{63}	-0.500	-0.512	0.478	0.433	0.433	0.956	-0.491	0.575	0.619	0.618	0.908
ρ	0.800	0.793	0.037	0.036	0.037	0.956	0.795	0.038	0.036	0.036	0.952
α	0.667	0.670	0.045	0.043	0.043	0.968	0.660	0.046	0.044	0.044	0.956

Table B.2: Posterior Summaries under the Nonignorable Model with a N(0, 1) Prior and a N(0, 10) Prior when missing percentage is high.

		N(0, 1) Prior					N(0, 10) Prior				
	TRUE	EST	SE	SIME	RMSE	CP	EST	SE	SIME	RMSE	CP
t=0											
β^*_{00}	-1.000	-1.036	0.143	0.128	0.132	0.992	-1.002	0.146	0.131	0.131	0.972
β^*_{10}	0.500	0.512	0.074	0.072	0.073	0.968	0.503	0.072	0.070	0.070	0.952
β^*_{20}	1.000	1.026	0.144	0.136	0.138	0.960	0.997	0.141	0.134	0.134	0.968
β^*_{30}	0.400	0.412	0.114	0.102	0.103	0.984	0.400	0.112	0.100	0.099	0.984
γ_{00}	-2.500	-2.568	0.240	0.234	0.243	0.960	-2.730	0.450	0.468	0.520	0.944
γ_{10}	0.500	0.511	0.118	0.107	0.107	0.984	0.499	0.129	0.118	0.118	0.976
γ_{20}	-0.500	-0.465	0.236	0.222	0.224	0.968	-0.492	0.257	0.244	0.244	0.960
γ_{30}	-0.500	-0.492	0.214	0.202	0.202	0.964	-0.502	0.218	0.205	0.205	0.960
γ_{60}	0.000	-0.106	0.589	0.467	0.478	0.984	0.049	0.991	0.876	0.876	0.968
t=1											
β^*_{01}	-1.000	-1.057	0.167	0.165	0.175	0.972	-0.977	0.183	0.185	0.186	0.928
β^*_{11}	0.500	0.498	0.077	0.074	0.074	0.972	0.499	0.077	0.074	0.074	0.964
β^*_{21}	1.000	1.022	0.152	0.150	0.151	0.948	0.970	0.154	0.149	0.152	0.924
β^*_{31}	-0.200	-0.193	0.112	0.106	0.106	0.956	-0.197	0.111	0.104	0.104	0.948
γ_{01}	-2.000	-2.008	0.214	0.183	0.182	0.956	-2.286	0.467	0.499	0.574	0.964
γ_{11}	0.500	0.523	0.089	0.092	0.094	0.924	0.500	0.097	0.101	0.101	0.924
γ_{21}	-0.500	-0.470	0.180	0.176	0.178	0.968	-0.529	0.199	0.201	0.202	0.936
γ_{31}	-0.250	-0.282	0.159	0.150	0.153	0.952	-0.256	0.166	0.156	0.156	0.964
γ_{41}	0.400	0.375	0.286	0.293	0.294	0.944	0.396	0.301	0.305	0.304	0.952

γ_{51}	-0.250	-0.157	0.267	0.240	0.257	0.952	-0.285	0.316	0.320	0.322	0.932
γ_{61}	0.500	0.228	0.653	0.535	0.599	0.972	0.684	1.077	1.090	1.104	0.932
t=2											
β^*_{02}	-1.000	-1.083	0.163	0.154	0.174	0.940	-1.006	0.167	0.166	0.166	0.928
β^*_{12}	0.500	0.491	0.077	0.073	0.073	0.956	0.495	0.076	0.071	0.071	0.960
β^*_{22}	1.000	1.051	0.156	0.148	0.156	0.960	0.996	0.155	0.149	0.149	0.956
β^*_{32}	-0.400	-0.408	0.119	0.114	0.114	0.956	-0.393	0.117	0.112	0.112	0.948
γ_{02}	-2.800	2.760	0.245	0.246	0.249	0.940	-3.002	0.367	0.444	0.487	0.924
γ_{12}	0.500	0.519	0.088	0.095	0.096	0.936	0.500	0.092	0.100	0.100	0.940
γ_{22}	-0.500	-0.461	0.182	0.163	0.167	0.968	-0.530	0.195	0.185	0.187	0.964
γ_{32}	0.250	0.237	0.161	0.171	0.171	0.920	0.270	0.168	0.181	0.182	0.916
γ_{42}	1.700	1.699	0.164	0.167	0.166	0.944	1.777	0.186	0.198	0.212	0.948
γ_{52}	-0.600	-0.495	0.270	0.284	0.302	0.924	-0.618	0.286	0.334	0.334	0.908
γ_{62}	1.300	0.983	0.546	0.507	0.597	0.912	1.426	0.717	0.841	0.848	0.892
t=3											
β^*_{03}	-1.000	-0.999	0.210	0.220	0.219	0.936	-0.972	0.221	0.243	0.244	0.904
β^*_{13}	0.500	0.516	0.102	0.098	0.100	0.976	0.505	0.101	0.101	0.101	0.944
β^*_{23}	1.000	1.019	0.176	0.168	0.168	0.956	0.987	0.174	0.163	0.164	0.952
β^*_{33}	-0.600	-0.609	0.155	0.159	0.159	0.944	-0.597	0.152	0.157	0.157	0.948
γ_{03}	-0.500	-0.543	0.127	0.131	0.138	0.936	-0.547	0.134	0.148	0.154	0.928
γ_{13}	0.500	0.502	0.062	0.063	0.062	0.956	0.503	0.064	0.067	0.067	0.952
γ_{23}	-0.500	-0.505	0.116	0.122	0.122	0.924	-0.504	0.119	0.127	0.127	0.932
γ_{33}	0.500	0.512	0.108	0.113	0.113	0.936	0.510	0.110	0.115	0.115	0.944
γ_{43}	1.700	1.738	0.135	0.140	0.144	0.948	1.733	0.138	0.142	0.146	0.956
γ_{53}	0.600	0.572	0.180	0.182	0.184	0.948	0.580	0.188	0.206	0.207	0.932
γ_{63}	-0.500	-0.440	0.405	0.416	0.419	0.948	-0.475	0.446	0.517	0.517	0.916
ρ	0.800	0.794	0.043	0.041	0.041	0.960	0.796	0.044	0.041	0.041	0.960
α	0.667	0.670	0.050	0.047	0.047	0.968	0.656	0.052	0.048	0.049	0.960

Bibliography

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.

Bekker, L.-G., Beyrer, C., and Quinn, T. C. (2012). Behavioral and biomedical combination strategies for hiv prevention. *Cold Spring Harbor perspectives in medicine*, **2**(8), a007435.

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., Shao, Q.-M., and Xu, D. (2002). Necessary and sufficient conditions on the properiety of posterior distributions for generalized linear mixed models. *Sankhyā: The Indian Journal of Statistics, Series A*, **64**, 57–85.

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.

Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2012). *Monte Carlo methods in Bayesian computation*. Springer Science & Business Media.

Chen, Q., Chen, M.-H., Ohlssen, D., and Ibrahim, J. G. (2013). Bayesian modeling and inference for clinical trials with partial retrieved data following dropout. *Statistics in Medicine*, **32**(24), 4180–4195.

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.

Genest, C. and Nešlehová, J. (2007). A primer on copulas for count data. *Astin Bulletin*, **37**(02), 475–515.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for gibbs sampling. *Applied Statistics*, pages 337–348.

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.

Kalichman, S. C., Cherry, C., Kalichman, M. O., Amaral, C. M., White, D., Pope, H., Swetzes, C., Eaton, L., Macy, R., and Cain, D. (2011). Integrated behavioral intervention to improve hiv/aids treatment adherence and reduce hiv transmission. *American Journal of Public Health*, **101**(3), 531–538.

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

Koblin, B. A., Bonner, S., Powell, B., Metralexis, P., Egan, J. E., Patterson, J., Xu, G., Hoover, D. R., Goodman, K., Chin, J., *et al.* (2012). A randomized trial of a behavioral intervention for black men who have sex with men: The dish study. *AIDS (London, England)*, **26**(4), 483.

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, C. (2004). Robit regression: a simple robust alternative to logistic and probit regression. *Applied Bayesian modeling and causal inference from incomplete-data perspectives*, pages 227–238.

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.

Starkweather, A. R., Lyon, D. E., Kinser, P., Heineman, A., Sturgill, J. L., Deng, X., Siangphoe, U., Elswick, R., Greenspan, J., and Dorsey, S. G. (2016). Comparison of low back pain recovery and persistence: A descriptive study of characteristics at pain onset. *Biological research for nursing*, **18**(4), 401–410.

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. *British Medical Journal*, **338**, b2393.

Stukel, T. A. (1988). Generalized logistic models. *Journal of the American Statistical Association*, **83**(402), 426–431.

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

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.

Zellner, A. (1986). On assessing prior distributions and bayesian regression analysis with g-prior distributions. *Bayesian inference and decision techniques: Essays in Honor of Bruno De Finetti*, **6**, 233–243.

Zhang, D., Chen, M.-H., Ibrahim, J. G., Boye, M. E., Wang, P., and Shen, W. (2014a). 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, D., Chen, M.-H., Ibrahim, J. G., Boye, M. E., and Shen, W. (2017). Bayesian model assessment in joint modeling of longitudinal and survival data with applications to cancer clinical trials. *Journal of Computational and Graphical Statistics*, **In press**.

Zhang, Y., Chen, M.-H., Ibrahim, J. G., Zeng, D., Chen, Q., Pan, Z., and Xue, X. (2014b). Bayesian gamma frailty models for survival data with semi-competing risks and treatment switching. *Lifetime Data Analysis*, **20**(1), 76–105.