JOINT MODELING OF MULTIVARIATE ORDINAL LONGITUDINAL OUTCOME

by

Zhen Jiang

B.S. in Biotechnology, Shanghai JiaoTong University, China, 2006

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Zhen Jiang

It was defended on

August 9th 2011

and approved by

Abdus S. Wahed, PhD, Associate Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Steven Belle, PhD, MScHyg, Professor, Department of Epidemiology, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Sati Mazumdar, PhD, Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Lan Kong, PhD, Assistant Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Richard A. Bilonick, PhD, Assistant Professor, Department of Ophthalmology, School of Medicine, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Dissertation Director: Abdus S. Wahed, PhD, Associate Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

JOINT MODELING OF MULTIVARIATE ORDINAL LONGITUDINAL OUTCOME

Zhen Jiang, PhD

University of Pittsburgh, 2011

Adherence to medication is critical to achieving effectiveness of any treatment. Poor adherence often results in lack of treatment effects, worsening of diseases and increased health care costs. Therefore, it has significant public health importance. However, determining factors that influence adherence behavior is complicated because adherence is often measured on multiple drugs over a long period of time, resulting in multivariate ordinal longitudinal outcome. In the first part of this dissertation, we present a joint model which assumes ordered outcomes arose from a partitioned latent multivariate normal process. This joint model provides a framework for analyzing multivariate ordered longitudinal data with a general multilevel association structure, covering both between and within outcome correlation within each individual. Simulation studies show that the estimators of regression parameters are more efficient than those obtained through fitting separate standard GEE for each outcome, though estimators from each method are unbiased. The proposed method also yields unbiased estimators for correlation parameters given the correct correlation structure. However, standard GEE estimators are biased when missing data are present and data are not missing completely at random (MCAR). In the second part of this dissertation, we apply inverse probability weighted (IPW) estimating equations to the proposed joint model to obtain consistent estimators when data are missing at random (MAR). Simulation studies show that IPW estimators are consistent when the missing model is correctly specified. Furthermore, we observe that fitting with correct correlation structures can also help reduce bias for standard GEE estimators. This demonstrates both a better correlation structure and a

better missing model will reduce bias in the analysis of missing at random longitudinal data using IPW GEE. We illustrate application of the proposed joint model to the Virahep-C data.

Keywords: Joint modeling; Generalized estimating equations; Inverse probability weighting; Latent variable model; Multivariate ordinal longitudinal data; Adherence to medication.

TABLE OF CONTENTS

LIST OF TABLES

PREFACE

My sincere thanks goes to my advisor Dr. Abdus S. Wahed for his mentoring and encouragement which has helped me through many difficulties for the past two years. I would also like to thank my dissertation committee Dr. Sati Mazumdar, Dr. Lan Kong, Dr. Steve Belle, and Dr. Richard Bilonick for their valuable comments and support.

Finally, I dedicate this work to my parents and my fiance for their endless love, support, and patience when it was needed the most.

1.0 INTRODUCTION

In the first part of this dissertation, we present a joint model for analysis of multivariate ordinal longitudinal outcome which assumes that the ordinal outcomes arose from a partitioned latent multivariate normal distribution. Generalized estimating equations are used to draw inferences on regression parameters and the least squares method is used to estimate the correlation parameters. In the second part of this dissertation, we apply inverse probability weighting (IPW) estimating equations to the proposed joint model to obtain consistent estimators when data are missing at random (MAR).

In Chapter 1, we introduce the Virahep-C study that motivated our research (Section 1.1), along with some relevant topics including medication adherence (Section 1.2), multivariate threshold model or latent multivariate normal distribution(Section 1.3), generalized estimating equations (Section 1.4), correlation structure (Section 1.5), methods for missing data (Section 1.6) and inverse probability weighted generalized estimating equations (Section 1.7). We plan to write Chapter 2 and Chapter 3 of this dissertation as independent papers. Therefore, there may be some redundancy in the introduction section.

1.1 VIRAHEP-C STUDY

This dissertation was motivated by the Virahep-C (Viral Resistance to Antiviral Therapy of Chronic Hepatitis C) study (Smith et al., 2007)[31], a nonrandomized, multicenter clinical trial designed to compare clinical response rates to peginterferon and ribavirin therapy between previously untreated African American and Caucasian American participants with chronic hepatitis C of genotype 1. The Virahep-C study enrolled 401 participants between

September 2002 and January 2004. Of these 401 participants, 196 (48.9%) are African American (AA) and 205 (51.1%) are Caucasian American (CA). All participants were to receive treatment with peginterferon alfa-2a weekly and ribavirin twice daily. The purpose of this dissertation is to identify potential influential factors for medication adherence of both medications. Investigating adherence behavior and potential influential factors is particularly important for subjects taking medication for hepatitis C virus (HCV) infection because a growing body of literature in HCV research has indicated that patients' treatment responses are affected by how closely prescribed medications were followed and how much medication was taken (Raptopoulou et al., 2005; Conjeevaram et al., 2006; Shiffman et al., 2007)[21, 5, 30]. Thus, identifying patients who are less likely to be adherent to their medication based on patient characteristics is critical, so that physicians can design early interventions to improve adherence in these patients.

In the Virahep-C study, medication adherence was measured by electronic monitors placed inside the caps of prescription bottles referred to as MEMS (Medication Event Management System, Aardex, Zug, Switzerland) caps. These electronic monitors continuously recorded an event any time that a bottle was closed, which was presumed to be the time a dose was taken. This information provided a detailed profile of each subject's adherence behavior. Based on the number of cap closings, daily adherence to ribavirin was categorized as fully adherent (2 closings), partially adherent (1 closing), or non-adherent (no closings) for each day, and weekly adherence to peginterferon was categorized as fully adherent (1 closing) or nonadherent (no closings) for each week. Thus, each subject's longitudinal adherence outcome consists of two components: one binary and the other ordinal.

1.2 MEDICATION ADHERENCE

Adherence to medication is defined as the extent to which patients follow their prescribed treatment regimens. Its measure can be dichotomous (i.e. adherence vs. nonadherence), ordinal or continuous percentage. Some available methods for measuring adherence include pill counts, self-report, monitoring drug concentration and electronic monitors (e.g. MEMS).

Adherence is critical for achieving effectiveness of any effective medical treatments. Poor adherence often results in lack of treatment effects, worsening of diseases and increased health care costs (Osterberg et al., 2005)[17]. Research has shown that subjects who do not follow their treatments have inferior prognoses compared to subjects who do (Horwitz et al., 1993; LaRosa 2000)[9, 11]. Unfortunately, poor adherence to medication is common even in well monitored clinical trials, especially in treating chronic diseases such as hypertension (Waeber et al., 1999)[35] or psychiatric illness (Nose et al., 2003)[16]. Race, gender and socioeconomic status have been found to be associated with adherence behavior in some studies[1, 32, 3].

Despite its critical impact, adherence and potential influential factors are difficult to investigate. Chronic disease patients often take multiple medications for their conditions over a long period of time, resulting in multivariate longitudinal adherence measures. Patients' adherence status cannot be fully characterized by a single adherence outcome and joint modeling of multiple outcomes is a natural choice. In addition, adherence to one medication is expected to be correlated with adherence to other medications, since a patient who is adherent to one medication is expected to be adherent to other medications as well. However, adverse effects of one medication might result in poor adherence to that medication only. Therefore, the structure of such correlation can be complex. This correlation between multiple adherence outcomes may be critical in modeling adherence as an outcome, and in modeling the association between treatment responses and adherence.

1.3 MULTIVARIATE THRESHOLD MODEL

Harville and Mee (1984), Qu et al. (1995) and Qu et al. (1992)[7, 19, 20] proposed a multivariate threshold model for analyzing clustered ordinal data which assumes the ordinal outcomes arose from a partitioned latent multivariate normal process. Suppose there are J ordinal responses $\boldsymbol{Y}=(Y_1, Y_2, \ldots, Y_J)$ each with K categories (i.e. $Y_i \in (1, 2, \ldots, K)$). The multivariate threshold model assumes that $\boldsymbol{Y}$ is a manifestation of a latent multivariate normal vector $\boldsymbol{u}=(u_1, u_2, \ldots, u_J)$ with mean zero and polychoric correlation $\boldsymbol{R}$ where $u_i \sim$

$N(0,1)$. More specifically, this model assumes $Y_i = k$ when the value of u_i falls into the kth of K partitioned intervals of latent normal distribution (i.e. $\Phi^{-1}(Pr(Y_i \leq k-1)) < u_i \leq \Phi^{-1}(Pr(Y_i \leq k))$) where Φ is the cdf of standard normal distribution. Then the pairwise joint cumulative probability and pairwise joint probability are given as follows:

$$Pr(Y_i \leq s, Y_j \leq t) = \int_{-\infty}^{\Phi^{-1}(Pr(Y_i \leq s))} \int_{-\infty}^{\Phi^{-1}(Pr(Y_j \leq t))} \phi_2(u_i, u_j, \rho_{i,j})\, \mathrm{d}u_i\, \mathrm{d}u_j, \tag{1.1}$$

$$Pr(Y_i = s, Y_j = t) = \int_{\Phi^{-1}(Pr(Y_i \leq s-1))}^{\Phi^{-1}(Pr(Y_i \leq s))} \int_{\Phi^{-1}(Pr(Y_j \leq t-1))}^{\Phi^{-1}(Pr(Y_j \leq t))} \phi_2(u_i, u_j, \rho_{i,j})\, \mathrm{d}u_i\, \mathrm{d}u_j, \tag{1.2}$$

where $i, j = (1, 2, \ldots, J)$, $s, t = (1, 2, \ldots, K)$, Φ is the cdf of standard normal distribution, ϕ_2 is the pdf of the bivariate normal distribution and $\rho_{i,j}$ is the (i,j)th elements of $\boldsymbol{R}$.

One advantage of the multivariate threshold model, as Qu et al. (1995)[19] argued, is that the assumption of a latent multivariate normal process does not affect the consistency of marginal regression estimates because the latent process does not affect the marginal probabilities. Rather, it only affects the correlation parameters through the pairwise joint probabilities. Thus even if the latent process is misspecified, marginal regression estimates remain unbiased, while correlation estimates can still be viewed as an approximation of the true correlation structure. In addition, the multivariate threshold models use polychoric correlation as an association measurement which is not restricted by the marginal probabilities as compared to the Pearson correlation; and, also the number of parameters in the polychoric correlation does not increase with the number of categories of the ordinal response increases when compared to the odds ratio. Finally, although Qu et al. (1995)[19] applied this method to clustered ordinal outcomes, their study was limited to univariate ordinal longitudinal data.

1.4 GENERALIZED ESTIMATING EQUATIONS (GEE)

Generalized estimating equations (GEE) proposed by Liang and Zeger (1986)[12] is perhaps the most widely used method for longitudinal data analysis. For each subject $i \in (1, 2, \ldots, N)$ with n_i repeated observations, let Y_i denotes an $n_i \times 1$ response vector, $V_i(\boldsymbol{\alpha})$ denotes a $n_i \times n_i$

covariance matrix where $\boldsymbol{\alpha}$ is correlation parameter, $\boldsymbol{\beta}$ denotes a $p \times 1$ parameter vector, X_i denotes a $p \times n_i$ covariate matrix, and u denotes a link function (e.g. logistic or log link). Thus, the generalized estimating equation is given as:

$$\sum_{i=1}^{N} \frac{\partial u(X_i^T \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} V_i(\boldsymbol{\alpha})^{-1} \{Y_i - u(X_i^T \boldsymbol{\beta})\} = 0 \tag{1.3}$$

GEE estimates can be obtained by an iterative Fisher scoring algorithm as follow:

$$\tilde{\boldsymbol{\beta}}^{(m+1)} = \tilde{\boldsymbol{\beta}}^{(m)} + \left(\sum_{i=1}^{N} \boldsymbol{D}_i^T(\tilde{\boldsymbol{\beta}}^{(m)}) V_i^{-1}(\boldsymbol{\alpha}) \boldsymbol{D}_i(\tilde{\boldsymbol{\beta}}^{(m)}) \right)^{-1} \sum_{i=1}^{N} \boldsymbol{D}_i^T(\tilde{\boldsymbol{\beta}}^{(m)}) V_i^{-1}(\boldsymbol{\alpha}) (\mathbf{Y_i} - u(X_i^T \tilde{\boldsymbol{\beta}}^{(m)}))) \tag{1.4}$$

where $\boldsymbol{D}_i^T(\boldsymbol{\beta}) = \partial u(X_i^T \boldsymbol{\beta}) / \partial \boldsymbol{\beta}^T$. As outlined in Liang and Zeger (1986)[12], the GEE estimate of $\boldsymbol{\beta}$ is consistent and follows an asymptotic multivariate normal distribution with mean $\boldsymbol{\beta}$ and variance covariance matrix $cov(\hat{\boldsymbol{\beta}})$, which can be consistently estimated by the robust variance estimate given as:

$$\hat{cov}(\hat{\boldsymbol{\beta}}) = \mathbf{A_n}^{-1} \mathbf{B_n} \mathbf{A_n}^{-1} \tag{1.5}$$

where $\mathbf{A_n} = \sum_{i=1}^{N} D_i(\hat{\boldsymbol{\beta}})^T V_i(\boldsymbol{\alpha})^{-1} D_i(\hat{\boldsymbol{\beta}})$ and $\mathbf{B_n} = \sum_{i=1}^{N} D_i(\hat{\boldsymbol{\beta}})^T V_i(\boldsymbol{\alpha})^{-1} (Y_i - u(X_i^T \hat{\boldsymbol{\beta}}))(Y_i - u(X_i^T \hat{\boldsymbol{\beta}}))^T V_i(\boldsymbol{\alpha})^{-1} D_i(\hat{\boldsymbol{\beta}})$. Unlike likelihood based methods, GEE requires specification of only the first two moments rather than the complete joint distribution of repeated observations. This enables it to handle various type of outcomes through different link functions. In GEE, correlation parameters are considered as nuisance parameters and consistent marginal regression estimates can be obtained even under a misspecified covariance structure. However, marginal regression estimates will be more efficient when the covariance structure is correctly specified.

Since the introduction of GEE, numerous improvements have been made to serve different purposes. For binary outcomes, Prentice (1988)[18] proposed to use correlation as an association measurement. Liang, Zeger and Qaqish (1992)[13] proposed to use the odds ratio as an association measurement, and Carey et al. (1993)[4] developed the alternating logistic regression, which resulted in efficient odds ratio estimators. Qu et al. (1992)[20] proposed a latent variable model for clustered binary outcomes where a tetrachoric correlation (the correlation of normally distributed latent processes for binary outcomes) is used as an association measurement. For ordinal outcomes, Miller et al. (1993)[15] presented GEE to handle univariate ordinal longitudinal data using a working correlation based on the inverse of Fisher's z transformation. Williamson et al. (1995) [36] proposed an estimating equations approach for clustered ordinal data using a global odds ratio as an association measurement. Qu et al. (1995)[19] extended their previous work for clustered binary outcomes in Qu et al. (1992) [20] and proposed a latent variable model for clustered ordinal outcomes where a polychoric correlation (the correlation of normally distributed latent processes for ordinal outcomes) is used as the association measurement.

Analysis of multivariate ordered longitudinal data where some components are binary and the others are ordinal is complex. The analysis requires a flexible model to handle the correlation between distinct outcomes at the same or different times and the correlation within the same outcome at different times. Sutradhar et al. (2000)[33] proposed a GEE method to analyze multivariate ordinal longitudinal data which accounted for both the structural correlation from the ordinal nature of the responses and longitudinal correlation from the repeated measurements over time.

All the above GEE based methods yield consistent marginal regression estimates when data are complete or missing completely at random(MCAR)(Rubin, 1976)[27]. But if the data are missing at random (MAR)(Rubin, 1976)[27], the estimates are biased. Robins et al. (1995)[24] proposed a weighted generalized estimating equations (WGEE) for longitudinal data with missing, which yields consistent estimators when data are MAR. Yi and Cook (2002)[37] proposed weighted second order estimating equations which facilitate consistent estimation of marginal regression parameters and association parameters. Lipsitz et al. (2009)[14] developed a joint GEE method for multivariate binary longitudinal outcomes with missing data, which yielded unbiased estimates when data are MAR.

1.5 CORRELATION STRUCTURE

In the analysis of longitudinal data using GEE, the correlation structure must be specified. There are several commonly used correlation structures including

(1) Independent:

$$\text{corr}=\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \tag{1.6}$$

(2) Compound symmetric:

$$\text{corr}=\begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{pmatrix} \tag{1.7}$$

(3) AR(1):

$$\text{corr}=\begin{pmatrix} 1 & \rho & \dots & \rho^{n-1} \\ \rho & 1 & \dots & \rho^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \dots & 1 \end{pmatrix} \tag{1.8}$$

(4) M-dependent:

$$\text{corr}=\begin{pmatrix} 1 & \rho_1 & \dots & \rho_{n-1} \\ \rho_1 & 1 & \dots & \rho_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \dots & 1 \end{pmatrix} \tag{1.9}$$

(5) Unstructured:

$$\text{corr}=\begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{12} & 1 & \dots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1n} & \rho_{2n} & \dots & 1 \end{pmatrix}. \tag{1.10}$$

The above correlation structures can only capture a single level of clustering. However, for data with more than one level of clustering (i.e. multilevel clustering), a more general form of correlation structure is needed. We will present, in chapter 2, a general multilevel correlation structure, which can be viewed as a natural extension of the above common correlation structures. More specifically, for the ith subject, let Y_{ijt} and $Y_{ij't'}$ be the jth outcome response observed at time t and the j'th outcome response observed at time t' respectively. There are two levels of clustering between Y_{ijt} and $Y_{ij't'}$: the first level is the association within the same outcome, the second level is the association between different outcomes. We can construct an extended AR-type correlation structure as follows:

$$\rho_{jt,j't'} = \alpha_{jj'}^{|t-t'|} \times \alpha_{2jj'}^{I(j \neq j')}, \tag{1.11}$$

where $-1 \leq \alpha_{jj'} \leq 1$ and $-1 \leq \alpha_{2jj'} \leq 1$. In the extended AR-type correlation, responses from the same outcome ($j = j'$) at different time points ($t \neq t'$) have an AR(1) correlation $\rho_{jt,j't'} = \alpha_{jj'}^{|t-t'|}$, responses from different outcomes ($j \neq j'$) at different time points ($t \neq t'$) have a weighted AR(1) correlation $\rho_{jt,j't'} = \alpha_{jj'}^{|t-t'|} \times \alpha_{2jj'}$ with weight $\alpha_{2jj'}$, and latent processes for the responses from different outcomes ($j \neq j'$) at the same time point ($t = t'$) have a correlation $\rho_{jt,j't'} = \alpha_{2jj'}$. Similarly, we can also construct an extended exchangeable type correlation and an extended m-dependent type correlation as follows:

$$\rho_{jt,j't'} = \alpha_{jj'} \times \alpha_{2jj'}^{I(j \neq j')}, \tag{1.12}$$

and $$\rho_{jt,j't'} = \alpha_{jj'|t-t'|} \times \alpha_{2jj'}^{I(j \neq j')}. \tag{1.13}$$

One can construct even more flexible correlation structures by combining different correlation structures at different levels. Correlation with more than two levels of clustering can be modeled in similar fashion.

1.6 METHODS FOR MISSING DATA

Missing data is a common phenomenon in longitudinal studies. The missing mechanism refers to how the probabilities of missingness relate to observed and unobserved data. Let R_{it} be the completeness indicator for Y_{it} where $R_{it} = 1$ if Y_{it} is observed and $R_{it} = 0$ if not observed. Rubin (1976)[27] and Laird (1988)[10] described three types of missing mechanisms as follows:

(1) Missing completely at random (MCAR): the probability of missing is not dependent on either observed or unobserved data such that: $\Pr(R_{it} = 1|Y_i, X_i)$=$\Pr(R_{it} = 1)$. When data are MCAR, the observed data can be considered as a random sample of the complete data. Therefore, all methods applicable to complete data can be applied to missing completely at random data.
(2) Missing at random (MAR): the probability of missing is dependent on previously observed responses but not the current response such that: $\Pr(R_{it} = 1|Y_i, X_i)$=$\Pr(R_{it} = 1|Y_{i1}, \ldots, Y_{it-1}, X_i)$.
(3) Missing not at random (NMAR): the missing mechanism is NMAR when neither MCAR or MAR condition holds.

Missing data in longitudinal studies will not only cause loss of information but also introduce potential bias in analysis. Therefore, missing data have significant implications in data analysis. General approaches for analysis of longitudinal missing data include:
(1) Complete cases analysis: incomplete observations are discarded and only subjects with complete data are included in the analysis. This method can lead to serious bias if data are not MCAR and loss of efficiency due to smaller sample size.
(2) Imputation-based procedures: missing data are filled in with imputed values to create a complete dataset for which standard methods can be applied. Common imputation methods include mean imputation, regression imputation, and multiple imputation (Rubin 1978)[28], which provide unbiased estimates if the imputation model is correctly specified.
(3) Model-based procedures: a model is defined for observed data and missing-data process. Likelihood based inferences are drawn based on this model. Common model-based methods include selection models and pattern-mixture models.
(4) Weighting procedures: this method assumes that the missing mechanism causes some responses to be unobserved with higher probabilities than others. Such underrepresented responses can be properly accounted for by weighting them using the inverse probabilities of observing those responses. First, the missing process is modeled to construct the probability of observing responses based on patient characteristics and time points. And then, responses are weighted by the inverse of the estimated observing probabilities in the estimating procedure to represent otherwise similar unobserved responses.

1.7 INVERSE PROBABILITY WEIGHTED GENERALIZED ESTIMATING EQUATIONS

The standard GEE method, as one of the most widely used methods for longitudinal data analysis, can fit unbalanced longitudinal data when some observations of subjects are missing. However, the consistency of standard GEE estimators depends on the missing mechanism. As stated by Liang and Zeger (1984)[12], standard GEE inferences are valid only when data are missing completely at random (MCAR). When data are MAR, there are two general approaches in obtaining consistent estimators under GEE framework: multiple imputation (Rubin, 1978 [28]) and inverse probability weighting (Robins et al. 1995 [24]). The idea of multiple imputation (MI) is to fill in the unobserved data with imputed data based on an assumed model. Then standard methods (e.g. GEE) can be applied to the imputed complete dataset.The inverse probability weighting (IPW) method was first introduced in survey studies by Horvitz and Thompson (1952) [8]. The general idea is that certain underlying missing mechanisms cause some responses to be observed with lower probabilities than others. To obtain consistent estimators, such underrepresented responses should be weighted by the inverse of observing probabilities. Robins et al. (1995) [24] proposed a class of IPW estimating equations which extended standard GEE method to MAR data and provided consistent marginal mean estimators when the response and missing models are correctly specified. In Robins et al. (1995) [24], the missing model is fitted using logistic regression:

$$\lambda_{it}(\gamma) = P(R_{it} = 1 | R_{it-1} = 1) = \frac{exp(w_i^T \gamma)}{1 + exp(w_i^T \gamma)} \tag{1.14}$$

$$\bar{\lambda}_{it}(\gamma) = P(R_{it} = 1) = \prod_{s=1}^{t} \lambda_{is}(\gamma) \tag{1.15}$$

where w_i is the covariate matrix for the missing model and γ is the missing parameter. The corresponding estimating equations for the missing parameter γ is given as:

$$\sum_{i=1}^{N} \frac{\partial \lambda_{\mathtt{i}}(\gamma)}{\partial \gamma^T} V_{\gamma i}^{-1} \{R_i - \lambda_i(\gamma)\} = 0, \tag{1.16}$$

where $\lambda_i(\gamma) = (\lambda_{i1}(\gamma), \lambda_{i2}(\gamma), \ldots, \lambda_{iT}(\gamma))^T$, $R_i = (R_{i1}, R_{i2}, \ldots, R_{iT})^T$ and $V_{\gamma i}^{-1} = [\lambda_i(\gamma)[1 - \lambda_i(\gamma)]^T]^{-1}$. Once the response probability of each response has been estimated, one can construct estimating equations for the mean parameter $\boldsymbol{\beta}$ by incorporating inverse probability as weights (IPW). The weighted estimating equations for marginal mean parameter is given as:

$$\sum_{i=1}^{N} \frac{\partial u(X_i^T \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} V_i(\boldsymbol{\alpha})^{-1} \Delta_{\beta i}(\hat{\gamma}) \{Y_i - u(X_i^T \boldsymbol{\beta})\} = 0 \tag{1.17}$$

where

$$\Delta_{\beta i}(\hat{\gamma}) = \begin{pmatrix} \Delta_{\beta i1}(\hat{\gamma}) & 0 & \ldots & 0 \\ 0 & \Delta_{\beta i2}(\hat{\gamma}) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \Delta_{\beta iT}(\hat{\gamma}) \end{pmatrix}, \tag{1.18}$$

$$\Delta_{\beta it}(\hat{\gamma}) = \frac{R_{it}}{\bar{\lambda}_{it}}. \tag{1.19}$$

Compared with the standard GEE equation (1.3) in section 1.4, the only extra term in the class of IPW estimating equation introduced by Robins et al. (1995)[24] is the weight matrix $\Delta_{\beta i}(\hat{\gamma})$. The idea here is that the underlying missing mechanism causes certain response Y_{it} to be observed with probability $\bar{\lambda}_{it}$. Such Y_{it} should be weighted by $R_{it}/\bar{\lambda}_{it}$ in estimating mean parameters to account for the unobserved outcome otherwise similar to Y_{it}.

Later, Yi and Cook (2002)[37] proposed inverse probability weighted second-order estimating equations for which the inverse probability weighting is applied to estimating equations for both mean and association parameters. They showed that, under certain assumptions, such methods provided consistent estimators for both mean and association parameters.

2.0 JOINT MODELING OF MULTIVARIATE ORDINAL LONGITUDINAL OUTCOMES

2.1 INTRODUCTION

Adherence to medication is defined as the extent to which patients follow their treatment regimens. Effectiveness of any treatment can be achieved only if patients take their medications as prescribed. Unfortunately, poor adherence to medication is common even in well monitored clinical trials, especially in treating chronic diseases such as hypertension (Waeber et al., 1999) and psychiatric illness (Nose et al., 2003). Poor adherence often results in lack of treatment effects, worsening of diseases and increased health care costs (Osterberg et al., 2005). Research has shown that subjects who do not follow their treatments have inferior prognosis than subjects who do (Horwitz et al., 1993; LaRosa 2000). Despite its critical impact, adherence behavior and potential influential factors are difficult to investigate because adherence is often measured on more than one medication repeatedly over a long period of time. In addition, adherence to one medication is expected to be correlated with adherence to other medications, since a patient who is adherent to one medication is expected to be adherent to other medications as well. However, adverse effects of one medication might result in poor adherence to that medication compared to others. Therefore, the structure of such correlation can be complex. This correlation between multiple adherence outcomes may be critical in modeling adherence as an outcome, and in modeling the association between treatment responses and adherence.

The dataset that motivated this research originates from the Virahep-C (Viral Resistance to Antiviral Therapy of Chronic Hepatitis C) study (Smith et al., 2007). It is a nonrandomized, multicenter clinical trial designed to compare clinical response rates to peginterferon and ribavirin therapy between previously untreated African American and Caucasian American participants with chronic hepatitis C of genotype 1. Investigating adherence behavior and potential influential factors is particularly important for subjects taking medication for the hepatitis C virus (HCV) infection because a growing body of literature in HCV research has indicated that patients' treatment responses are affected by how closely prescribed medications were followed and how much medication was taken (Raptopoulou et al., 2005; Conjeevaram et al., 2006). Thus, identifying patients who are less likely to be adherent to their medication based on patient characteristics is critical, so that early interventions can be implemented to improve adherence in these patients.

In the Virahep-C study, the initial prescription was peginterferon alfa-2a (180 mcg/wk) weekly and ribavirin(1000-1200 mg/day) twice daily. Adherence was measured through electronic monitors placed inside the caps of prescription bottles. These caps are referred as MEMS (Medication Event Management System, Aardex, Zug, Switzerland) caps. These monitors continuously recorded an event any time that a bottle was closed, which was presumed to be the time a dose was taken. This information provided a detailed profile of each subject's adherence behavior. Based on the number of cap closing, adherence to ribavirin was categorized as fully adherent (2 closing), partially adherent (1 closing), or non-adherent (no closing) for each day, and adherence to peginterferon was categorized as fully adherent (1 closing) or nonadherent (no closing) for each week. Thus, each subject's longitudinal adherence outcomes consist of two components: a binary longitudinal outcome and an ordinal longitudinal outcome.

Generalized estimating equations (GEE) proposed by Liang and Zeger (1986) is perhaps the most widely used method for longitudinal data analysis. Since its introduction, numerous improvements have been made to GEE to serve different purposes. For binary outcomes, Prentice (1988) proposed to use correlation as association measurement. Liang, Zeger and Qaqish (1992) proposed to use the odds ratio as association measurement, and Carey et al. (1993) developed the alternating logistic regression, which resulted in efficient odds ratio estimators. Qu et al. (1992) proposed a latent variable model for clustered binary outcomes where a tetrachoric correlation (the correlation of normally distributed latent processes for binary outcomes) is used as association measurement. For ordinal outcomes, Miller et al. (1993) presented GEE to handle univariate ordinal longitudinal data using a working correlation based on the inverse of Fisher's z transformation. Williamson et al. (1995) proposed an estimating equations approach for clustered ordinal data using a global odds ratio as association measurement. Qu et al. (1995) extended their previous work for clustered binary outcomes in Qu et al. (1992) and proposed a latent variable model for clustered ordinal outcomes where a polychoric correlation (the correlation of normally distributed latent processes for ordinal outcomes) is used as the association measurement. An advantage of polychoric correlation is that the number of correlation parameters, unlike the odds ratio, does not increase as the number of categories in each outcome increases. Although Qu et al. (1995) applied this method to clustered ordinal outcomes, their study was limited to univariate ordinal longitudinal data.

Analysis of multivariate ordered longitudinal data where some components are binary and the others are ordinal is complex. The analysis requires a flexible model to handle the correlation between distinct outcomes at the same or different times and the correlation within the same outcome at different times. Sutradhar et al. (2000) proposed a GEE method to analyze multivariate ordinal longitudinal data which accounted for both the structural correlation from the ordinal nature of the responses and longitudinal correlation from the repeated measurements over time. Lipsitz et al.(2009) developed a joint GEE method for multivariate binary longitudinal outcomes with missing data, which yielded almost unbiased estimates when data are missing at random (MAR) (Rubin, 1976).

In this chapter, we propose a joint model to analyze multivariate ordered longitudinal outcome where it is assumed that the observed multivariate ordered outcome is from a partitioned latent multivariate normal distribution. The GEE approach is used to draw inference on regression parameters and the least square method is used to estimate correlation parameters. This proposed joint model provides a flexible framework to account for the multilevel correlation structure covering both between and within outcome associations. Furthermore, simulation studies show that the GEE estimators of regression parameters obtained from the joint model are unbiased, and more efficient than those obtained from fitting separate GEE for each outcome.

As a brief overview of this chapter, we first describe the proposed joint model (section 2.2) followed by how to draw inference using GEE (section 2.3). Then a series of simulation results are presented to examine the joint model (section 2.4) followed by the application to Virahep-C data (section 2.5). Finally, we discuss the implications and limitations of the proposed joint model (section 2.6).

2.2 THE JOINT MODEL

Suppose that for each subject i ($i = 1, \dots, n$), there are J ordered longitudinal outcomes each with K_j categories ($j = 1, \dots, J$) and the jth outcome of ith subject has n_{ij} repeated measurements observed at time t ($t = T_{j1}, \dots, T_{jn_{ij}}$). Denote Y_{ijt} as the jth ordinal outcome of ith subject observed at time t where Y_{ijt} takes value from the set $\{0, \dots, K_j - 1\}$ and Y_{ijt} is binary if $K_j = 2$. It follows that the aggregated response vector $\boldsymbol{Y_i}$ of subject i can be formed as $\boldsymbol{Y_i = (Y'_{i1}, \dots, Y'_{iJ})'}$ where $\boldsymbol{Y_{ij}}$=$(Y_{ijT_{j1}}, \dots, Y_{ijT_{jn_{ij}}})'$ is the response vector for jth outcome in subject i.

2.2.1 MARGINAL PROBABILITY MODEL

Let $\gamma_{ijtk} = Pr\left(Y_{ijt} \leq k\right)$ and $\pi_{ijtk} = Pr\left(Y_{ijt} = k\right)$ $(k = 0, \dots, K_j - 2)$ be the marginal cumulative probability and probability of Y_{ijt} at the kth category where γ_{ijtk} is assumed to depend upon covariates through a cumulative logistic regression model. Furthermore, we denote $\mathbf{X_{it}} = (x_{it1}, \dots, x_{itp})$ as a $p \times 1$ dimensional covariate vector for subject i at time t, which may include both subject and time specific covariates. We also denote $\boldsymbol{\beta}_j = (\beta_{j1}, \beta_{j2}, \dots, \beta_{jp})'$ as the $p \times 1$ dimensional regression coefficient vector of $\mathbf{X_{it}}$ for jth outcome and $\boldsymbol{a}_j = (a_{j0}, a_{j1}, \dots, a_{jK_j-2})'$ as the $(K_j - 1)$ dimensional intercept vector for jth outcome with K_j categories, such that $a_{j0} < a_{j1} < \dots < a_{jK_j-2}$. Then the cumulative probability γ_{ijtk} and probability π_{ijtk} of Y_{ijt} can be given as:

$$\gamma_{ijtk} = Pr\left(Y_{ijt} \leq k \mid \mathbf{X_{it}}, \boldsymbol{\beta}_j, \boldsymbol{a}_j\right) = \frac{exp(a_{jk} + \mathbf{X_{it}^T}\boldsymbol{\beta}_j)}{1 + exp(a_{jk} + \mathbf{X_{it}^T}\boldsymbol{\beta}_j)}, \tag{2.1}$$

$$\pi_{ijtk} = Pr\left(Y_{ijt} = k \mid \mathbf{X_{it}}, \boldsymbol{\beta}_j, \boldsymbol{a}_j\right) = \frac{exp(a_{jk} + \mathbf{X_{it}^T}\boldsymbol{\beta}_j)}{1 + exp(a_{jk} + \mathbf{X_{it}^T}\boldsymbol{\beta}_j)} - \frac{exp(a_{jk-1} + \mathbf{X_{it}^T}\boldsymbol{\beta}_j)}{1 + exp(a_{jk-1} + \mathbf{X_{it}^T}\boldsymbol{\beta}_j)}. \quad (2.2)$$

When Y_{ijt} is binary (i.e. $K_j = 2$), equations (2.1) and (2.2) are reduced to modeling failure probability using the following logistic regression model:

$$\gamma_{ijt0} = Pr\left(Y_{ijt} \leq 0 \mid \mathbf{X_{it}}, \boldsymbol{\beta}_j, \boldsymbol{a}_j\right) = \frac{exp(a_{j0} + \mathbf{X_{it}^T}\boldsymbol{\beta}_j)}{1 + exp(a_{j0} + \mathbf{X_{it}^T}\boldsymbol{\beta}_j)}, \quad (2.3)$$

$$\pi_{ijt0} = Pr\left(Y_{ijt} = 0 \mid \mathbf{X_{it}}, \boldsymbol{\beta}_j, \boldsymbol{a}_j\right) = \gamma_{ijt0}. \quad (2.4)$$

Although in theory each outcome can have its own set of covariates, for the purpose of simplicity we assume that each outcome has the same set of covariates. We allow distinct regression coefficients $\boldsymbol{\beta}_j$ for each outcome. When no ambiguity exists, we will use $\boldsymbol{\beta} = (\boldsymbol{a}'_1, \boldsymbol{\beta}'_1, \boldsymbol{a}'_2, \boldsymbol{\beta}'_2, \ldots, \boldsymbol{a}'_J, \boldsymbol{\beta}'_J)'$ to denote the overall parameter vector.

2.2.2 JOINT PROBABILITY MODEL

To jointly model the multivariate longitudinal outcome $\boldsymbol{Y_i} = \boldsymbol{(Y'_{i1}, \ldots, Y'_{iJ})'}$ where $\boldsymbol{Y_{ij}} = (Y_{ijT_{j1}}, \ldots, Y_{ijT_{jn_{ij}}})'$, in this section, we construct the pairwise joint probability for any two responses in $\boldsymbol{Y_i}$. We define $\gamma_{i,jtk,j't'k'} = Pr\left(Y_{ijt} \leq k, Y_{ij't'} \leq k'\right)$ and $\pi_{i,jtk,j't'k'} = Pr(Y_{ijt} = k, Y_{ij't'} = k')$ as the joint cumulative probability and joint probability respectively, of responses Y_{ijt} and $Y_{ij't'}$. We assume that Y_{ijt} and $Y_{ij't'}$ originate from a bivariate normal distribution $(u, v) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{jt,j't'} \\ \rho_{jt,j't'} & 1 \end{pmatrix}\right)$ partitioned by threshold values $\Phi^{-1}(\gamma_{ijtk})$ and $\Phi^{-1}(\gamma_{ij't'k'})$. Thus, $\gamma_{i,jtk,j't'k'}$ and $\pi_{i,jtk,j't'k'}$ can be given as:

$$\gamma_{i,jtk,j't'k'} = Pr(Y_{ijt} \leq k, Y_{ij't'} \leq k') = \int_{-\infty}^{\Phi^{-1}(\gamma_{ijtk})} \int_{-\infty}^{\Phi^{-1}(\gamma_{ij't'k'})} \phi_2(u, v, \rho_{jt,j't'})\, \mathrm{d}u\, \mathrm{d}v, \quad (2.5)$$

$$\pi_{i,jtk,j't'k'} = Pr(Y_{ijt} = k, Y_{ij't'} = k') = \int_{\Phi^{-1}(\gamma_{ijtk-1})}^{\Phi^{-1}(\gamma_{ijtk})} \int_{\Phi^{-1}(\gamma_{ij't'k'-1})}^{\Phi^{-1}(\gamma_{ij't'k'})} \phi_2(u, v, \rho_{jt,j't'})\, \mathrm{d}u\, \mathrm{d}v, \quad (2.6)$$

where Φ is the cdf of a standard normal distribution and ϕ_2 is the pdf of a bivariate normal distribution with pairwise correlation $\rho_{jt,j't'}$ which will be constructed in section 2.3. Notice that the threshold values $\Phi^{-1}(\gamma_{ijtk})$ and $\Phi^{-1}(\gamma_{ijtk})$ are determined by the marginal probability model and only the joint probability model is affected by the pairwise correlation $\rho_{jt,j't'}$.

This method of modeling joint cumulative probabilities can be viewed as a multivariate threshold model (Harville and Mee, 1984; Qu et al., 1992; Qu et al., 1995)[7, 20, 19]. It is assumed that the $\sum_{j=1}^{J} n_{ij} \times 1$ dimensional response vector $\boldsymbol{Y_i}$ is observed from partitioning a $\sum_{j=1}^{J} n_{ij} \times 1$ dimensional latent random vector $\boldsymbol{\varepsilon}_i = (\boldsymbol{\varepsilon}_{i1}, \ldots, \boldsymbol{\varepsilon}_{iJ})'$ where $\boldsymbol{\varepsilon}_{ij} = (\varepsilon_{ijT_{j1}}, \ldots, \varepsilon_{ijT_{jn_{ij}}})'$. $\boldsymbol{\varepsilon}_i$ is assumed to follow a multivariate normal distribution with mean zero and correlation matrix $\mathbf{R}$ where $\rho_{jt,j't'}$ is one of the elements in $\mathbf{R}$. It is further assumed that one will observe $y_{ijt} = k$ when $\Phi^{-1}(\gamma_{ijtk-1}) < \varepsilon_{ijt} \leq \Phi^{-1}(\gamma_{ijtk})$. Thus, the multivariate threshold method ensures the marginal cumulative probability of $Y_{ijt} \leq k$ is still $\Phi\left(\Phi^{-1}(\gamma_{ijtk})\right) = \gamma_{ijtk}$, and the joint cumulative probability of $Y_{ijt} \leq k$ and $Y_{ij't'} \leq k'$ is given in equation (2.5).

2.2.3 CORRELATION STRUCTURE

As demonstrated in section 2.2, given the marginal probabilities, the joint probability of any two response variables can be calculated by equation (2.6) as long as the correlation matrix $\mathbf{R}$ of latent random vector $\boldsymbol{\varepsilon}_i$ is specified. There are two levels of associations in $\mathbf{R}$: the first level is the association within the outcome at different time points; the second level is the association between different outcomes at the same or different time points. To accommodate this feature in $\mathbf{R}$, we develop a general multilevel correlation structure similar to the correlation model proposed by Lipsitz et al. (2009)[14]. It can be viewed as a natural extension of common univariate correlation structures (e.g. AR, exchangeable and m-dependent). For example, we construct an extended AR-type correlation for ε_{ijt} and $\varepsilon_{ij't'}$ as:

$$\rho_{jt,j't'} = \alpha_{jj'}^{|t-t'|} \times \alpha_{2jj'}^{I(j \neq j')}, \tag{2.7}$$

where $-1 \leq \alpha_{jj'} \leq 1$ and $-1 \leq \alpha_{2jj'} \leq 1$. In the extended AR-type correlation structure, latent processes for the responses from the same outcome $(j = j')$ at different time points $(t \neq t')$ have an AR(1) correlation $\rho_{jt,j't'} = \alpha_{jj'}^{|t-t'|}$, latent processes for the responses from different outcomes $(j \neq j')$ at different time points $(t \neq t')$ have a weighted AR(1) correlation $\rho_{jt,j't'} = \alpha_{jj'}^{|t-t'|} \times \alpha_{2jj'}$ with weight $\alpha_{2jj'}$, and latent processes for the responses from different outcomes $(j \neq j')$ at same time point $(t = t')$ have a correlation $\rho_{jt,j't'} = \alpha_{2jj'}$. Similarly, we can also construct an extended exchangeable correlation and an extended m-dependent correlation as follows:

$$\rho_{jt,j't'} = \alpha_{jj'} \times \alpha_{2jj'}^{I(j \neq j')}, \tag{2.8}$$

and
$$\rho_{jt,j't'} = \alpha_{jj'|t-t'|} \times \alpha_{2jj'}^{I(j \neq j')}. \tag{2.9}$$

One can construct even more flexible correlation structures by combining different correlation structures at different levels. For the extended AR-type and exchangeable correlation model defined as equations (2.7) and (2.8), only J^2 and $\frac{1}{2}(J^2 + J)$ correlation parameters are needed for J longitudinal outcomes. More parameters are required for the extended m-dependent structure in equation (2.9), in which case the estimation of correlation parameter $\boldsymbol{\alpha}$ may be computationally intensive.

2.3 INFERENCE

In this section, we demonstrate how to draw inference on regression parameters from the proposed joint model using GEE. We first construct subject specific variance-covariance matrix which accounts for the structural correlation due to the ordinal nature of the responses along with both between and within outcome correlation.

2.3.1 COVARIANCE STRUCTURE

To construct the variance-covariance matrix, let us first dichotomize ordinal variable Y_{ijt} with K_j categories into a $(K_j - 1)$ dimensional binary vector Z_{ijt} where Z_{ijt} and its expectation $\boldsymbol{\pi}_{ijt}$ =E(Z_{ijt}) are given as:

$$Z_{ijt} = (z_{ijt0}, z_{ijt1}, \ldots, z_{ijtK_j-2}), \tag{2.10}$$

$$\text{and} \qquad \boldsymbol{\pi}_{ijt} = (\pi_{ijt0}, \pi_{ijt1}, \ldots, \pi_{ijtK_j-2}), \tag{2.11}$$

where $z_{ijtk} = \begin{cases} 1 & \text{if } Y_{ijt} = k \\ 0 & \text{if } Y_{ijt} \neq k \end{cases}$, and $\pi_{ijtk} = Pr\left(Y_{ijt} = k \mid \mathbf{X_{it}}, \boldsymbol{\beta}_j, \boldsymbol{a}_j\right)$ as in equation (2.2), $k = \{0, \ldots, K_j - 2\}$. It follows that the $\sum_{j=1}^{J} n_{ij}$ dimensional ordinal response vector $\mathbf{Y_i}$ can be transformed into a $\sum_{j=1}^{J} n_{ij}(K_j - 1)$ dimensional binary vector $\mathbf{Z_i}$ with expectation $\boldsymbol{\pi}_i$=E$(\mathbf{Z_i})$.

The variance-covariance matrix $\boldsymbol{\Sigma}_i$ of the dichotomized binary response vector $\mathbf{Z_i}$ for ith subject can be written as:

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \Sigma_{i11} & \Sigma_{i12} & \ldots & \Sigma_{i1J} \\ \Sigma_{i21} & \Sigma_{i22} & \ldots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{iJ1} & \Sigma_{iJ2} & \ldots & \Sigma_{iJJ} \end{pmatrix}, \tag{2.12}$$

$$\text{where} \quad \Sigma_{ijj'} = \begin{pmatrix} \Sigma_{i,jT_{j1},j'T_{j'1}} & \Sigma_{i,jT_{j1},j'T_{j'2}} & \ldots & \Sigma_{i,jT_{j1},j'T_{j'n_{ij'}}} \\ \Sigma_{i,jT_{j2},j'T_{j'1}} & \Sigma_{i,jT_{j2},j'T_{j'2}} & \ldots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{i,jT_{jn_{ij}},j'T_{j'1}} & \Sigma_{i,jT_{jn_{ij}},j'T_{j'2}} & \ldots & \Sigma_{i,jT_{jn_{ij}},j'T_{n_{ij'}}} \end{pmatrix}, \tag{2.13}$$

where $\Sigma_{ijj'}$ is a $n_{ij}(K_j - 1) \times n_{ij'}(K'_j - 1)$ dimensional variance-covariance matrix for binary vectors $\mathbf{Z_{ij}}$ and $\mathbf{Z_{ij'}}$, and $\Sigma_{ijt,j't'}$ is a $(K_j - 1) \times (K'_j - 1)$ dimensional variance-covariance matrix for binary vectors Z_{ijt} and $Z_{ij't'}$. When $j = j'$ and $t = t'$, $\Sigma_{ijt,j't'}$ represents the structural correlation due to the polytomous nature of ordinal outcome Y_{ijt}. Otherwise, $j \neq j'$ or $t \neq t'$,

$\Sigma_{ijt,j't'}$ represents the correlation from the latent multivariate normal process for Y_{ijt} and $Y_{ij't'}$. The (k,k')th element of covariance matrix $\Sigma_{ijt,j't'}$ can be derived as:

$$\sigma_{i,jtk,j't'k'} = \begin{cases} -\pi_{ijtk}\pi_{ij't'k'} & k \neq k' \\ \pi_{ijtk}(1-\pi_{ij't'k'}) & k = k' \end{cases} \quad j = j' \quad \text{and} \quad t = t', \tag{2.14}$$

$$\sigma_{i,jtk,j't'k'} = \pi_{i,jtk,j't'k'} - \pi_{ijtk}\pi_{ij't'k'} j \neq j' \quad \text{or} \quad t \neq t', \tag{2.15}$$

where π_{ijtk} and $\pi_{ij't'k'}$ are marginal probabilities given by equation (2.2) and $\pi_{i,jtk,j't'k'}$ is the pairwise joint probability given by equation (2.5).

2.3.2 ITERATIVE ESTIMATION PROCEDURE

As the previous section demonstrated, the variance-covariance matrix $\mathbf{\Sigma}_i$ is a function of regression parameter $\boldsymbol{\beta}$ and correlation parameter $\boldsymbol{\alpha}$ through marginal probability π_{ijtk} and pairwise joint probability $\pi_{i,jtk,j't'k'}$ respectively. Because both $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are unknown, to estimate $\boldsymbol{\beta}$ using GEE, we resort to an iterative estimation process.

Had the correlation parameter $\boldsymbol{\alpha}$ been known, the regression parameter $\boldsymbol{\beta}$ could have been estimated by solving the following generalized estimating equation based on dichotomized binary response vector $\mathbf{Z_i}$:

$$\sum_{i=1}^{N} \boldsymbol{D}_i^T(\boldsymbol{\beta})\mathbf{\Sigma}_i^{-1}(\boldsymbol{\alpha},\boldsymbol{\beta})\{\mathbf{Z_i} - \boldsymbol{\pi}_i(\boldsymbol{\beta})\} = 0 \tag{2.16}$$

where $\boldsymbol{\pi}_i(\boldsymbol{\beta})$ as a function of $\boldsymbol{\beta}$ is defined in equation (2.2), $\mathbf{\Sigma}_i^{-1}(\boldsymbol{\alpha},\boldsymbol{\beta})$ as a function of both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is defined in equation (2.12) - equation (2.15), and $\boldsymbol{D}_i(\boldsymbol{\beta})$ is the partial derivative of $\boldsymbol{\pi}_i(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ whose element is given as follow:

$$\frac{\delta}{\delta\boldsymbol{\beta}_j}\boldsymbol{\pi}_{ijtk}(\boldsymbol{\beta}) = \begin{cases} \mathbf{X_{it}^T}\frac{exp(a_{jk}+\mathbf{X_{it}^T}\boldsymbol{\beta}_j)}{(1+exp(a_{jk}+\mathbf{X_{it}^T}\boldsymbol{\beta}_j))^2} & k=0, \\ \mathbf{X_{it}^T}\left[\frac{exp(a_{jk}+\mathbf{X_{it}^T}\boldsymbol{\beta}_j)}{(1+exp(a_{jk}+\mathbf{X_{it}^T}\boldsymbol{\beta}_j))^2} - \frac{exp(a_{jk-1}+\mathbf{X_{it}^T}\boldsymbol{\beta}_j)}{(1+exp(a_{jk-1}+\mathbf{X_{it}^T}\boldsymbol{\beta}_j))^2}\right] & k \geq 1. \end{cases} \tag{2.17}$$

We can solve equation (2.16) using an iterative equation:

$$\tilde{\boldsymbol{\beta}}^{(m+1)} = \tilde{\boldsymbol{\beta}}^{(m)} + \left(\sum_{i=1}^{N} \boldsymbol{D}_i^T(\tilde{\boldsymbol{\beta}}^{(m)})\mathbf{\Sigma}_i^{-1}(\boldsymbol{\alpha},\tilde{\boldsymbol{\beta}}^{(m)})\boldsymbol{D}_i(\tilde{\boldsymbol{\beta}}^{(m)})\right)^{-1} \sum_{i=1}^{N} \boldsymbol{D}_i^T(\tilde{\boldsymbol{\beta}}^{(m)})\mathbf{\Sigma}_i^{-1}(\boldsymbol{\alpha},\tilde{\boldsymbol{\beta}}^{(m)})(\mathbf{Z_i} - \boldsymbol{\pi}_i(\tilde{\boldsymbol{\beta}}^{(m)})) \tag{2.18}$$

However, $\boldsymbol{\alpha}$ is unknown and needs to be estimated. To estimate $\boldsymbol{\alpha}$, we first define $\mathbf{S_i}(\boldsymbol{\beta}) = (\mathbf{Z_i} - \boldsymbol{\pi}_i(\boldsymbol{\beta}))(\mathbf{Z_i} - \boldsymbol{\pi}_i(\boldsymbol{\beta}))'$. In addition, let $\mathbf{s_i}(\boldsymbol{\beta}) = vech(\mathbf{S_i}(\boldsymbol{\beta}))$ (Vonesh and Chinchilli, 1997)[34] be the vectorzied version of the upper diagonal elements in $\mathbf{S_i}(\boldsymbol{\beta})$ and $\boldsymbol{\sigma}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) = vech(\boldsymbol{\Sigma}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}))$ be the counterpart of $\boldsymbol{\Sigma}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})$.

Given E($\mathbf{S_i}(\boldsymbol{\beta})$)=$\boldsymbol{\Sigma}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})$ under the joint model, one can view $\mathbf{S_i}(\boldsymbol{\beta})$ as the "observed outcome" and $\boldsymbol{\Sigma}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})$ as its expectation to construct a regression model to estimate $\boldsymbol{\alpha}$ when $\boldsymbol{\beta}$ is known by minimizing the sum of the Euclidean norms of $(\mathbf{s}_i(\boldsymbol{\beta}) - \boldsymbol{\sigma}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}))$ as follow:

$$\sum_{i=1}^{N} (\mathbf{s}_i(\boldsymbol{\beta}) - \boldsymbol{\sigma}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}))^T (\mathbf{s}_i(\boldsymbol{\beta}) - \boldsymbol{\sigma}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})) \tag{2.19}$$

with respect to $\boldsymbol{\alpha}$. This is also equivalent to solving the following generalized estimating equation for $\boldsymbol{\alpha}$ when $\boldsymbol{\beta}$ is given:

$$\sum_{i=1}^{N} \frac{\partial \boldsymbol{\sigma}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \boldsymbol{\alpha}}^T \{\mathbf{s}_i(\boldsymbol{\beta}) - \boldsymbol{\sigma}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})\} = 0, \tag{2.20}$$

where an identity variance-covariance matrix is assumed for $\boldsymbol{\sigma}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})$.

Now we can carry out the iterative estimation procedure in the following manner:

Step 1: Obtain an initial estimate of $\boldsymbol{\beta}$ denoted by $\hat{\boldsymbol{\beta}}^0$, which can be done by fitting separate model to each outcome;

Step 2: Solve for $\hat{\boldsymbol{\alpha}}^0$ by minimizing equation (2.19) with $\boldsymbol{\beta}$ replaced by $\hat{\boldsymbol{\beta}}^0$;

Step 3: Solve for $\hat{\boldsymbol{\beta}}^1$ from generalized estimating equation (2.16) with $\boldsymbol{\alpha}$ replaced by $\hat{\boldsymbol{\alpha}}^0$.

Step 4: Iterate between Step 2 and 3 until convergence criteria are fulfilled for both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, where the solution is denoted by $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$.

$\hat{\boldsymbol{\beta}}$ obtained from above iterative procedure is consistent and follows an asymptotic normal distribution with mean $\boldsymbol{\beta}$ and a variance-covariance matrix that can be estimated by the robust variance estimator proposed by Liang and Zeger (1986)[12] as follow:

$$c\hat{o}v(\hat{\boldsymbol{\beta}}) = \mathbf{A_n}^{-1}\mathbf{B_n}\mathbf{A_n}^{-1} \tag{2.21}$$

where $\mathbf{A_n} = \sum_{i=1}^{N} D_i(\hat{\boldsymbol{\beta}})^T \Sigma_i(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})^{-1} D_i(\hat{\boldsymbol{\beta}})$ and $\mathbf{B_n} = \sum_{i=1}^{N} D_i(\hat{\boldsymbol{\beta}})^T \Sigma_i(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})^{-1} (\mathbf{Z_i} - \boldsymbol{\pi}_i(\hat{\boldsymbol{\beta}}))(\mathbf{Z_i} - \boldsymbol{\pi}_i(\hat{\boldsymbol{\beta}}))^T \Sigma_i(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})^{-1} D_i(\hat{\boldsymbol{\beta}})$.

2.4 SIMULATION STUDY

To examine the proposed joint model, we performed a series of simulation studies. For simplicity, we assumed each subject i had one ordinal longitudinal outcome $\boldsymbol{Y_{i1}}$ with three categories and one binary longitudinal outcome $\boldsymbol{Y_{i2}}$. We constructed 5 repeated measurements for the ordinal outcome at times $t = 0, 1, 2, 3, 4$ as $\boldsymbol{Y_{i1}}$=$(Y_{i10}, Y_{i11}, Y_{i12}, Y_{i13}, Y_{i14})'$ and 3 repeated measurements for the binary outcome at times $t' = 0, 2, 4$ as $\boldsymbol{Y_{i2}}$=$(Y_{i20}, Y_{i22}, Y_{i24})'$. Thus, each subject had a 8×1 dimensional response vector $\boldsymbol{Y_i = (Y'_{i1}, Y'_{i2})'}$. Similar to the Virahep-C data, this is a case where two outcomes are measured at different sets of times by design.

We used the cumulative logistic regression to specify the true marginal probabilities for the ordinal outcome, and used the logistic model to specify the true marginal probabilities for the binary outcome. For each subject, we constructed one time covariate T specified as above and one baseline covariate X generated from a Bernoulli(0.5) distribution. We first specified the regression coefficient $\boldsymbol{\beta}_1 = (\beta_{1x}, \beta_{1t})$ and the intercept $\boldsymbol{a}_1 = (a_{10}, a_{11})$ for the ordinal outcome and $\boldsymbol{\beta}_2 = (\beta_{2x}, \beta_{2t})$ and $\boldsymbol{a}_2 = a_{20}$ for the binary outcome. Then, we constructed the marginal cumulative probabilities for both outcomes $\boldsymbol{\gamma}_i = (\boldsymbol{\gamma}_{i1}, \boldsymbol{\gamma}_{i2})$ by equations (2.1) and (2.3) where $\boldsymbol{\gamma}_{i1} = (\boldsymbol{\gamma}_a i10, \boldsymbol{\gamma}_{i11}, \boldsymbol{\gamma}_{i12}, \boldsymbol{\gamma}_{i13}, \boldsymbol{\gamma}_{i14})$ and $\boldsymbol{\gamma}_{i2} = (\boldsymbol{\gamma}_{i20}, \boldsymbol{\gamma}_{i22}, \boldsymbol{\gamma}_{i24})$.

To construct the latent multivariate normal process for both outcomes, we generated a 8×1 random vector $\boldsymbol{\epsilon_i}$=$(\epsilon_{i10}, \epsilon_{i11}, \epsilon_{i12}, \epsilon_{i13}, \epsilon_{i14}, \epsilon_{i20}, \epsilon_{i22}, \epsilon_{i24})$. Each element of $\boldsymbol{\epsilon_i}$ followed a standard normal distribution and $\boldsymbol{\epsilon_i}$ followed a multivariate normal distribution with correlation $\mathbf{R}^*$ which is a function of the correlation parameters $\boldsymbol{\alpha} = (\alpha_{11}, \alpha_{22}, \alpha_{12}, \alpha_{212})$. In this simulation study, $\mathbf{R}^*$ was constructed as either an extended AR-type correlation by equation (2.7) or an extended exchangeable correlation by equation (2.8). After obtaining the marginal probabilities and latent process $\boldsymbol{\epsilon_i}$ with correlation $\mathbf{R}^*$, each element of response vector $\boldsymbol{Y_i = (Y'_{i1}, Y'_{i2})'}$ was generated as follows:

$$Y_{i1t} = \begin{cases} 0, \text{ if } 0 < \Phi(\epsilon_{i1t}) \leq \gamma_{i1t0}, \\ 1, \text{ if } \gamma_{i1t0} < \Phi(\epsilon_{i1t}) \leq \gamma_{i1t1}, \\ 2, \text{ if } \gamma_{i1t1} < \Phi(\epsilon_{i1t}) < 1, \qquad t \in \{0, 1, 2, 3, 4\}, \end{cases}$$

$$Y_{i2t'} = \begin{cases} 0, & \text{if } 0 < \Phi(\epsilon_{i2t'}) \le \gamma_{i2t'0}, \\ 1, & \text{if } \gamma_{i2t'0} < \Phi(\epsilon_{i2t'}) < 1, \qquad t' \in \{0, 2, 4\}. \end{cases}$$

The above process ensures that the generated response vector $\boldsymbol{Y_i}$ has specified marginal probabilities given by equations (2.1) and (2.3) and a specified correlation for latent normal process given by either equation (2.7) or equation (2.8).

We examined different correlation structures and parameters. In each scenario, both $M = 500$ Monte-Carlo datasets with $n = 100$ subjects and $M = 1000$ Monte-Carlo datasets with $n = 50$ subjects were generated. We fit four models to each dataset:
(1) sep-GLM: separately fit cumulative logistic regression for ordinal outcome and logistic regression for binary outcome using maximum likelihood method;
(2) sep-GEE: separately fit the ordinal outcome and the binary outcome using GEE;
(3) Joint GEE Independence: GEE applied to the joint model with independent correlation structure;
(4) Joint GEE: GEE applied to the joint model where correlation parameter $\boldsymbol{\alpha}$ is estimated along with regression parameter $\boldsymbol{\beta} = (\boldsymbol{a}_1', \boldsymbol{\beta}_1', \boldsymbol{a}_2', \boldsymbol{\beta}_2')'$.

Tables 1 and 2 show the simulation results based on different sample sizes n=100 and n=50. In each table, we examined six sets of $\boldsymbol{\alpha}$ representing strong, moderate and no correlation for both extended AR-type correlation and extended exchangeable correlation. Regression coefficients were estimated by all four methods, and Monte-Carlo variances were calculated based on M replications. For sep-GLM, model based variance estimates were calculated while for sep-GEE and both Joint GEE methods with or without independent assumption, robust sandwich variance estimates were calculated.

The results in Table 1 (n=100) show that the estimates of the regression coefficients are approximately unbiased across all four models. Notice that even in scenarios 1, 2, 4, and 5 where the true correlation is not zero, fitting sep-GEE and Joint GEE with independent assumption still yielded unbiased estimates. This is expected because in theory, GEE estimators remain unbiased even when the correlation is misspecified (Liang and Zeger 1986). In sep-GLM, the model-based variances grossly underestimated the true variance, leading to poor coverage probabilities for the 95% confidence interval (CI), sometimes as low as 67%. On the other hand, in all GEE based methods, the coverage probabilities of the 95%

CI match the nominal level and the robust variance estimates match the Monte-Carlo variances, implying that robust variance estimators are approximately unbiased even when the correlation is misspecified.

Results in scenarios 3 and 6 of Table 1 indicate that when the two longitudinal outcomes are uncorrelated, separate or joint modeling makes no difference in efficiency as demonstrated by the approximately equal MC variances under each method. However, when the two processes are correlated and the joint model is adopted with correct correlation structure, the estimators gain efficiency and in some cases, by as much as 23%. The estimates of baseline covariates for the binary outcome ($\hat{\beta}_{2x}$) generally have the most efficiency gain, and the efficiency gain also increased as the correlation increased where the efficiency gain of $\hat{\beta}_{2x}$ is 7% and 3% for scenarios 2 and 5, and 23% and 13% for scenarios 1 and 4. The estimates of baseline covariates for the ordinal outcome ($\hat{\beta}_{1x}$) have slight efficiency gain up to 7% as seen in scenario 1. The estimates of time covariates for both outcomes ($\hat{\beta}_{1t}, \hat{\beta}_{2t}$) have efficiency gain up to 6% and 10% as seen in scenario 1. Table 2, with sample size n=50, shows very similar results as in Table 1, except that the standard errors of estimates were larger compared to those presented in Table 1 due to the decreased sample size.

In addition, the results in Table 3 show the joint model yielded unbiased estimates for correlation parameters given the correct correlation structure, which may be of further interest when assessing the association among multiple outcomes. Finally, as a check of the robustness of our model when the correlation structure is misspecified, we generated data from both extended exchangeable and AR-type correlations. In both cases, we fitted the joint model with three working correlations, namely, independent, extended AR-type and extended exchangeable correlation. The results in Table 4 show that our model still yielded unbiased estimates with coverage probabilities close to nominal level in all cases even when the correlation structure is misspecified, but the variance of estimates is minimized when the working correlation structure matches the true correlation.

2.5 DATA ANALYSIS

A growing body of literature in hepatitis C virus research has shown that adherence to medications affects treatment responses (Raptopoulou et al., 2005; Conjeevaram et al., 2006; Shiffman et al., 2007)[21, 5, 30]. Our main goal in this study is to identify potential determining factors for adherence to medication in Virahep-C patients. The Virahep-C study enrolled 401 participants between September 2002 and January 2004. Of these 401 participants, 196 (48.9%) are African American (AA) and 205 (51.1%) are Caucasian American (CA). Initial treatment strategy was peginterferon alfa-2a (180 mcg/wk) weekly and ribavirin(1000-1200 mg/day) twice daily. Although therapy continued for up to 48 weeks, it is believed that adherence during the first 12 weeks is more important to achieve a response to therapy than adherence after 12 weeks (Ferenci et. al 2005 [6]). Table 5 shows that less than 6% of the participants dropped out during the first 12 weeks after treatment initiates. In this analysis, we assumed drop outs were missing completely at random (MCAR) (Rubin, 1976; Laird, 1988)[27, 10]. Because it is only when the electronic monitors are malfunctioned that some responses would not be observed. Adherence to ribavirin was categorized into three levels (2, fully adherent; 1, partially adherent; 0, nonadherent), and adherence to peginterferon was categorized into two levels (1, adherent; 0, nonadherent).

Two adherence outcomes were modeled jointly with a common set of covariates. The main covariate of interest is race (1, CA; 0, AA). Other covariates include days (i.e. number of days since the treatment was initiated), days and race interaction, gender (1, male; 0, female), baseline HCV RNA level ($log_{10}\ eq/ml$), and employment status (1, unemployed; 0, employed). The marginal probabilities of adherence to ribavirin were assumed to depend upon covariates through the cumulative logistic regression model as follow:

$$log\left(\frac{Pr(Y_{i1t} \leq k)}{1 - Pr(Y_{i1t} \leq k)}\right) = a_{1k} + \beta_1 \times race_i + \beta_2 \times t + \beta_3 \times race_i * t + \beta_4 \times sex_i$$

$$+\beta_5 \times vloadblg_i + \beta_6 \times employ_i,$$

where t=0, 1, 2,…, 84 and k=0, 1. In this cumulative logistic model, a positive regression coefficient indicates a higher probability of being in a lower adherence category for ribavirin.

Similarly, the marginal probabilities of adherence to peginterferon were assumed to depend upon covariates through the logistic regression model as follow:

$$log\left(\frac{Pr(Y_{i2t} \leq 0)}{1 - Pr(Y_{i2t} \leq 0)}\right) = a_{20} + \beta_1' \times race_i + \beta_2' \times t + \beta_3' \times race_i * t + \beta_4' \times sex_i$$

$$+\beta_5' \times vloadblg_i + \beta_6' \times employ_i,$$

where t=0, 7, 14,...,84. In this logistic model, a positive regression coefficient indicates a higher probability of being nonadherent to peginterferon.

Although the true association structure of the Virahep-C data is unknown, it is unlikely that no association exists between two adherence outcomes from the same patient. The joint model also yields positive estimates for both between and within outcomes correlations assuming extended AR-type structure $(\hat{\alpha}_{11}, \hat{\alpha}_{22}, \hat{\alpha}_{12}, \hat{\alpha}_{22}) = (0.203, 0.981, 0.988, 0.417)$. Because adherence outcomes were observed repeatedly over time, it is reasonable to assume an extended AR-type correlation.

We fit both sep-GEE and the proposed joint model assuming an extended AR-type correlation to the Virahep-C data. Table 6 shows the regression parameter estimates from both models. Based on the results from the joint model, the main covariate of interest race is not significantly associated with neither of the two outcomes. Employment status is the only patient characteristic that has a statistically significant effect on the adherence to peginterferon. The odds ratio of being nonadherent to peginterferon is $exp(0.812) = 2.25$ for a patient who is unemployed at baseline compared to an otherwise similar patient who is employed at baseline. Adherence to ribavirin varied significantly over time. The odds ratio of being in a lower adherence category for ribavirin at any given day compared to the previous day is $exp(0.014) = 1.01$, controlling for other patient characteristics. The coefficients of time for both outcomes are positive indicating patients have a higher odds of being less adherent to their medications as time increases. In contrast, the results from sep-GEE in Table 6 indicates that race and days are statistically significant for both outcomes, while employment status is significant for the adherence to peginterferon only.

2.6 DISCUSSION

In this chapter, we proposed a joint model to analyze multivariate ordered longitudinal data where it is assumed that the ordered outcomes are observed from a partitioned latent normal process. The joint model accounts for both the between and within outcome association using the correlation of the latent multivariate normal process. This correlation ranges from -1 to 1 and is not restricted by marginal probabilities. The simulation study indicates that the GEE estimators of the joint model are unbiased and more efficient compared to those obtained from fitting separate GEE for each outcome. Furthermore, as observed in the results from the analysis of Virahep-C data, failure to appropriately account for the correlation between multiple outcomes can lead to improper conclusion.

In the joint model, although we made an assumption that observed ordered outcomes arose from a latent multivariate normal process, this assumption, as argued in Qu et al. (1995)[19], does not affect the unbiasness of $\hat{\boldsymbol{\beta}}$ because the latent normal process does not affect the marginal probabilities. Rather, it only affects the correlation parameters through the pairwise joint probabilities. Thus even if the latent process is misspecified, $\hat{\boldsymbol{\beta}}$ will remain unbiased, while $\hat{\boldsymbol{\alpha}}$ can still be viewed as an approximation of the true correlation structure.

We demonstrated in section 2.3 that the joint model can incorporate a general multilevel correlation structure, for example, extended AR-type, extended exchangeable, and extended m-dependent. One can construct even a more flexible correlation structure by combining different correlation structures. For example, one can assume latent processes for the responses from the same outcome at different time points have an AR(1) correlation structure $\rho_{jt,jt'} = \alpha_{jj}^{|t-t'|}$, while latent processes for the responses from different outcomes at different time points have either an extended exchangeable correlation structure $\rho_{jt,j't'} = \alpha_{2jj'}$ or an extended m-dependent correlation structure $\rho_{jt,j't'} = \alpha_{2jj'|t-t'|}$. These two extended mixed correlation structures can be written as:

$$\rho_{jt,j't'} = \alpha_{jj'}^{|t-t'| \times I(j=j')} \times \alpha_{2jj'}^{I(j \neq j')},$$

$$\text{and } \rho_{jt,j't'} = \alpha_{jj'}^{|t-t'| \times I(j=j')} \times \alpha_{2jj'|t-t'|}^{I(j \neq j')}.$$

The proposed joint model can also be extended to handle more levels of clustering (e.g. clustering due to sites, clinical centers etc.). Although additional correlation parameters are required, the estimation process will remain similar. Furthermore, because the joint model uses the least square method to estimate correlation parameter $\boldsymbol{\alpha}$, the model is able to provide consistent estimates, and more importantly, is applicable for all of the multilevel correlation structures mentioned above, whereas Liang and Zeger (1986) only provided formulas to estimate $\boldsymbol{\alpha}$ for several simple correlation structures (e.g. exchangeable, AR, m-dependent).

Finally, inference of the proposed joint model is based all available data. It yields consistent regression estimates ($\hat{\boldsymbol{\beta}}$) when data are missing completely at random (MCAR) (Rubin, 1976; Laird, 1988)[27, 10]. The proposed estimator will be biased when data are missing at random (MAR). In chapter 3, we apply inverse probability weighted estimating equations (Robins et al., 1995[24]) to the proposed joint model to obtain consistent estimators when data are missing at random (MAR).

3.0 JOINT MODELING OF MULTIVARIATE ORDINAL LONGITUDINAL OUTCOMES WITH MISSING DATA

In chapter 2, we presented a joint model with standard GEE inference for joint analysis of multivariate ordinal longitudinal outcome. As stated in Liang and Zeger (1984)[12], standard GEE inference is valid only when data are complete or missing completely at random (MCAR). Therefore, the proposed estimators are biased if data are missing at random. In this chapter, we applied the standard GEE with inverse probability weighting to obtain consistent estimating equations in the presence of missing at random data.

3.1 INTRODUCTION

Although the main purpose of designing longitudinal studies is to collect data from participants at every follow-up visit, despite the best effort from project personnel, missing observations are almost inevitable in practice. Such missing observations may have significant implications on the results of data analysis. Loss of information due to unobserved data will generally reduce the precision of the parameter estimates. However, the potential bias introduced by missingness in longitudinal studies is a much greater problem. For example, in the case that the reason for missing responses is treatment-related adverse events or lack of efficacy, a missing observation actually contains information related to the outcome of interest (e.g. treatment efficacy). In such cases, ignoring or inappropriately handling missing data might result in misleading conclusions.

The standard GEE method, as one of the most widely used methods for longitudinal data analysis, has the capacity to fit unbalanced longitudinal data when missing observations are present. In other words, one can draw standard GEE inference, even if some subjects only have measurements for a subset of all time points (i.e. available case analysis). However, the consistency of such standard GEE estimator depends on the relationship between the probability of missing and the outcome being modeled. Rubin (1976)[27] and Laird (1988)[10] described three types of such relationships, also called missing mechanism, as follows: (1) Missing completely at random (MCAR): the probability of missing depends on neither observed nor unobserved responses. (2) Missing at random (MAR): the probability of missing depends on previously observed responses but not the missing response. (3) Missing not at random (NMAR): neither MCAR nor MAR conditions hold and thus, the probability of missing depends on the missing response itself. When data are MCAR, the observed data can be considered as a random sample of the complete data. Therefore, unbiased inference can be drawn without modifying currently available standard methods for longitudinal studies. On the other hand, when data are MAR or NMAR, most statistical methods, including GEE provide biased inference. As stated in Liang and Zeger (1984)[12], standard GEE inference is valid only when data are MCAR.

In the case of MAR, there are two general approaches to obtaining consistent estimators under the GEE framework: multiple imputation (Rubin, 1978 [28]) and inverse probability weighting (Robins et al. 1995 [24]). The idea of multiple imputation (MI) is to fill in the unobserved data with imputed data based on an assumed model. Then standard methods (e.g. GEE) can be applied to the imputed complete dataset.

The inverse probability weighting (IPW) method was first introduced in survey studies by Horvitz and Thompson (1952) [8]. The general idea is that the underlying missing mechanism causes some responses to be observed with lower probabilities than others. In order to obtain consistent estimators, such underrepresented responses should be properly weighted by the inverse probabilities of being observed. First, missing process is modeled to construct the probability of observing responses. And then, in the estimating procedure, responses are weighted by the inverse of the estimated probablities of being observed to account for otherwise similar unobserved responses. Robins et al. (1995) [24] proposed a class of IPW estimating equations which extended the standard GEE method to MAR data. Their IPW estimator provides consistent estimators for marginal mean parameters when data are MAR. Later, Yi and Cook (2002)[37] proposed a weighted second-order estimating equations, where the inverse probability weighting was applied to the estimating equations for both mean and association parameters. They showed that, under certain assumptions, their method provides consistent estimators for both marginal mean parameters and association parameters.

In the case of NMAR, Rotnitzky and Robins (1997) [25] and Robins, Rotnitzky and Scharfstein (2000) [22] proposed methods to extend IPW estimators to the NMAR setting, which is beyond the scope of this dissertation.

In chapter 2, a joint model along with the inference procedure under GEE framework was presented for joint analysis of multiple ordinal longitudinal outcomes. In this chapter, we will apply inverse probability weighted estimating equations to draw inference of the proposed joint model when data are MAR. As a brief overview of this chapter, we first describe the response and missing process of the joint model in section 3.2 followed by the IPW inference procedure (section 3.3). Then a series of simulation results are shown to demonstrate how consistent estimators for mean parameters are obtained in the presence of MAR data. We compare these results to complete case (CC) analysis and available case (AC) analysis under different true correlation structures.

3.2 JOINT MODEL WITH MISSING DATA

3.2.1 THE RESPONSE PROCESS

The detailed response process of the proposed joint model is described in section 2.2. For the purpose of continuity in this section, only two key aspects are summarized: marginal probability model and joint probability model.

Suppose that for ith subject ($i = 1, \ldots, n$), there are J ordinal longitudinal outcomes each with K_j ($j = 1, \ldots, J$) categories. Denote Y_{ijt} as the jth outcome of ith subject at time t ($t = T_{j1}, \ldots, T_{jn_{ij}}$), where Y_{ijt} takes value from the set $\{0, \ldots, K_j - 1\}$. Let $\mathbf{X_{it}} = (x_{it1}, \ldots, x_{itp})$ be a $p \times 1$ covariate vector for subject i at time t. As stated in chapter

2, we model the marginal probability of ordinal outcome using cumulative logistic regression as follows:

$$\gamma_{ijtk} = Pr\left(Y_{ijt} \leq k \mid \mathbf{X_{it}}, \boldsymbol{\beta}_j, \boldsymbol{a}_j\right) = \frac{exp(a_{jk} + \mathbf{X_{it}^T}\boldsymbol{\beta}_j)}{1 + exp(a_{jk} + \mathbf{X_{it}^T}\boldsymbol{\beta}_j)}, \tag{3.1}$$

$$\pi_{ijtk} = Pr\left(Y_{ijt} = k \mid \mathbf{X_{it}}, \boldsymbol{\beta}_j, \boldsymbol{a}_j\right) = \frac{exp(a_{jk} + \mathbf{X_{it}^T}\boldsymbol{\beta}_j)}{1 + exp(a_{jk} + \mathbf{X_{it}^T}\boldsymbol{\beta}_j)} - \frac{exp(a_{jk-1} + \mathbf{X_{it}^T}\boldsymbol{\beta}_j)}{1 + exp(a_{jk-1} + \mathbf{X_{it}^T}\boldsymbol{\beta}_j)}, \tag{3.2}$$

where γ_{ijtk} $(k = 0, \ldots, K_j - 2)$ is the marginal cumulative probability of $Y_{ijt} \leq k$, π_{ijtk} is the marginal probability of $Y_{ijt} = k$, $\boldsymbol{\beta}_j = (\beta_{j1}, \beta_{j2}, \ldots, \beta_{jp})'$ is a $p \times 1$ regression coefficient vector for jth outcome and $\boldsymbol{a}_j = (a_{j0}, a_{j1}, \ldots, a_{jK_j-2})'$ is a $(K_j - 1)$ dimensional intercept vector for jth outcome with K_j categories, such that $a_{j0} < a_{j1} < \ldots < a_{jK_j-2}$.

To construct the pairwise joint probability, we assume that any pairs of Y_{ijt} and $Y_{ij't'}$ originate from a bivariate normal distribution $(u, v) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{jt,j't'} \\ \rho_{jt,j't'} & 1 \end{pmatrix}\right)$ partitioned by threshold values. The pairwise joint probability of responses Y_{ijt} and $Y_{ij't'}$ can be derived as:

$$\pi_{i,jtk,j't'k'} = Pr(Y_{ijt} = k, Y_{ij't'} = k') = \int_{\Phi^{-1}(\gamma_{ijtk-1})}^{\Phi^{-1}(\gamma_{ijtk})} \int_{\Phi^{-1}(\gamma_{ij't'k'-1})}^{\Phi^{-1}(\gamma_{ij't'k'})} \phi_2(u, v, \rho_{jt,j't'})\, \mathrm{d}u\, \mathrm{d}v, \tag{3.3}$$

where Φ is the cdf of a standard normal distribution and ϕ_2 is the pdf of a bivariate normal distribution with correlation $\rho_{jt,j't'}$. The correlation structure was constructed in section 2.3.

3.2.2 THE MISSING PROCESS

In this chapter, we only focus on dealing with missing at random (MAR) data, where the probability of missing depends on the previously observed responses rather than the missing response itself. Let R_{ijt} be an indicator such that $R_{ijt} = 1$ if Y_{ijt} is observed and $R_{ijt} = 0$ if Y_{ijt} is missing. We assume monotone missing (i.e. $R_{ijt} = 0$ implies $R_{ijt'} = 0\ \forall\ t' > t$) and also first measurements of all outcomes are observed (i.e. $R_{ij1} = 1\ \forall i, j$). Denote $\mathbf{H_{ijt}} = (Y_{ij1}, Y_{ij2}, \ldots, Y_{ijt-1})$ to be the history of jth outcome of ith subject up to time $t-1$.

The conditional probability of observing Y_{ijt} given the previous outcome Y_{ijt-1} is observed can be written as:

$$\lambda_{ijt} = Pr(R_{ijt} = 1|R_{ijt-1} = 1, \mathbf{Y_i}, \mathbf{X_i}, \gamma_j) = Pr(R_{ijt} = 1|R_{ijt-1} = 1, \mathbf{H_{ijt}}, \mathbf{X_i}, \gamma_j), \qquad (3.4)$$

and then the probability of observing Y_{ijt} can be written as:

$$\bar{\lambda}_{ijt} = Pr(R_{ijt} = 1|\mathbf{Y_i}, \mathbf{X_i}, \gamma_j) = Pr(R_{ijt} = 1|\mathbf{H_{ijt}}, \mathbf{X_i}, \gamma_j) = \prod_{s=1}^{t} \lambda_{ijs} \qquad (3.5)$$

where γ_j is outcome-specific missing parameters for jth outcome. Typically, logistic regression is used to model the missing process as follows:

$$logit(\lambda_{ijt}) = \gamma_{0j} + H_{ijt}^T \gamma_{1j} + X_{it}^T \gamma_{2j} \ (t > 1). \qquad (3.6)$$

For jth outcome, the missing parameter $\gamma_j = (\gamma_{0j}, \gamma_{1j}^T, \gamma_{2j}^T)$. When no ambiguity exists, we will use $\gamma = (\gamma_1', \ldots, , \gamma_J')'$ to denote the missing parameters. Note that when t=1, $\lambda_{ijt} = 1$.

We further assume that the probability of missing for jth outcome does not depend on other outcomes. That is,

$$Pr(R_{ijt} = 1|R_{ijt-1} = 1, R_{ij't'} = 1, \mathbf{Y_i}, \mathbf{X_i}, \gamma_j) = Pr(R_{ijt} = 1|R_{ijt-1} = 1, \mathbf{H_{ijt}}, \mathbf{X_i}, \gamma_j), \quad (3.7)$$

where $(j' \neq j)$. Therefore, one can derive that the conditional joint probability of observing a pair of $(Y_{ijt}, Y_{ij't'})$ given (Y_{ijt-1}, Y_{ijt-1}) equals the product of two marginal conditional probability of observing Y_{ijt} and $Y_{ij't'}$. That is,

$$\lambda_{ijt,ij't'}(\gamma) = P(R_{ijt} = 1, R_{ij't'} = 1|R_{ijt-1} = 1, R_{ij't'-1} = 1, \mathbf{Y_i}, \mathbf{X_i}) = \lambda_{ijt} \times \lambda_{ij't'}, \qquad (3.8)$$

where $t > t' - 1$ and $t' > t - 1$.

Then the probability of observing a pair of $(Y_{ijt}, Y_{ij't'})$ can be written as function of $\lambda_{ijt}(\gamma_j)$:

$$\bar{\lambda}_{i,jt,j't'}(\gamma) = P(R_{ijt} = 1, R_{ij't'} = 1|\mathbf{Y_i}, \mathbf{X_i}, \gamma) \begin{cases} \prod_{s=1}^{\max(t,t')} \lambda_{ijs}(\gamma_j), & \text{if } j = j', \\ \prod_{s=1}^{t} \lambda_{ijs}(\gamma_j) \prod_{s'=1}^{t'} \lambda_{ij's'}(\gamma_j'), & \text{if } j \neq j'. \end{cases}$$

3.3 INVERSE PROBABILITY WEIGHTED ESTIMATING EQUATIONS AND INFERENCE

In section 2.3, we demonstrated how to draw inference for the proposed joint model based on standard GEE. Robins et al. 1995 [24] along with Yi and Cook (2002)[37] extended standard GEE to handle missing at random data. They proposed to use inverse the probability of observing the outcome as weights in estimating equations. The response probabilities are modeled as a function of covariates and previously observed outcomes.

In this section, we apply inverse probability weighted estimating equation to draw inference for the proposed joint model. In section 3.3.1, we introduce the estimating equations for missing parameters under MAR assumption. In sections 3.3.2 and 3.3.3, we present the weighted estimating equations for both mean and association parameters of proposed joint model.

3.3.1 ESTIMATING EQUATIONS AND INFERENCE FOR MISSING PARAMETERS

Let us denote $\lambda_i(\gamma) = (\lambda'_{i1}, \ldots, \lambda'_{iJ})'$, λ_{ij}=$(\lambda_{ijT_{j1}}, \ldots, \lambda_{ijT_{jn_{ij}}})'$ where λ_{ijt} is defined in equation (3.4) and is a function of γ_j. Let $R_i = (R'_{i1}, \ldots, R'_{iJ})'$, R_{ij}=$(R_{ijT_{j1}}, \ldots, R_{ijT_{jn_{ij}}})'$ where R_{ijt} is the response indicator for Y_{ijt}. Because the missing process is modeled by logistic regression in section 3.2.2, the corresponding estimating equations for missing parameter γ is given as:

$$\sum_{i=1}^{N} W_i(\gamma) = \sum_{i=1}^{N} \frac{\partial \lambda_i(\gamma)}{\partial \gamma^T} V_{\gamma i}^{-1}\{R_i - \lambda_i(\gamma)\} = 0, \tag{3.9}$$

where $V_{\gamma i}^{-1} = [\lambda_i(\gamma)(1-\lambda_i(\gamma))^T]^{-1}$. We can solve γ using iterative algorithm:

$$\tilde{\gamma}^{(m+1)} = \tilde{\gamma}^{(m)} + \left(\sum_{i=1}^{N} \frac{\partial \lambda_i(\gamma)}{\partial \gamma^T} V_{\gamma i}^{-1} \frac{\partial \lambda_i(\gamma)}{\partial \gamma} \right)^{-1} \sum_{i=1}^{N} \frac{\partial \lambda_i(\gamma)}{\partial \gamma^T} V_{\gamma i}^{-1}\{R_i - \lambda_i(\gamma)\} \tag{3.10}$$

By simple algebra, we can simplify equation (3.9) as follow:

$$\sum_{i=1}^{N} W_i(\gamma) = \sum_{i=1}^{N} w_i^T\{R_i - \lambda_{\mathbf{i}}(\gamma)\} = 0, \tag{3.11}$$

where $w_i = (w_{i1}, \ldots, w_{iJ})$, $w_{ij} = (w_{ijT_{j1}}, \ldots, w_{ijT_{jn_{ij}}})$ and $w_{ijt} = (1, H_{ijt}, X_{it})$ is the design matrix in missing model (3.6). Iterative equation (3.10) can also be simplified as follow:

$$\tilde{\gamma}^{(m+1)} = \tilde{\gamma}^{(m)} + \left(\sum_{i=1}^{N} w_i^T V_{\gamma i} w_i\right)^{-1} \sum_{i=1}^{N} w_i^T \{R_i - \lambda_i(\gamma)\} \tag{3.12}$$

Above simplification is applicable to all estimating equations for logistic regression.

3.3.2 ESTIMATING EQUATIONS AND INFERENCE FOR MEAN PARAMETERS

Once the response probability of each outcome has been estimated, we can construct inverse probability weighted estimating equations for mean parameter $\boldsymbol{\beta}$ as follows:

$$\sum_{i=1}^{N} U_i(\alpha, \beta, \hat{\gamma}) = \sum_{i=1}^{N} D_i(\beta)^T V_{\beta i}^{-1} \Delta_{\beta i}(\hat{\gamma}) \{\mathbf{Z_i} - \boldsymbol{\pi}_i(\boldsymbol{\beta})\} = 0, \tag{3.13}$$

where

$$\Delta_{\beta i}(\hat{\gamma}) = \begin{pmatrix} \Delta_{\beta i1}(\hat{\gamma}) & 0 & \ldots & 0 \\ 0 & \Delta_{\beta i2}(\hat{\gamma}) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \Delta_{\beta iJ}(\hat{\gamma}) \end{pmatrix}, \tag{3.14}$$

$$\text{and} \quad \Delta_{\beta ij}(\hat{\gamma}) = diag(\bar{\lambda}_{ij1}(\hat{\gamma})^{-1} R_{ij1}, \bar{\lambda}_{ij2}(\hat{\gamma})^{-1} R_{ij2}, \ldots, \bar{\lambda}_{ijT_{nj}}(\hat{\gamma})^{-1} R_{ijT_{nj}}). \tag{3.15}$$

Compared to standard GEE estimating equation (2.16) in section 2.3.2, the only extra term here is the weight matrix $\Delta_{\beta i}(\hat{\gamma})$. This class of IPW GEE estimating equation was introduced by Robins et al. (1995). The idea is that the underlying missing mechanism causes response Y_{ijt} to be observed with probability $\bar{\lambda}_{ijt}(\hat{\gamma})$. Such Y_{ijt} should be weighted by $\bar{\lambda}_{ijt}(\hat{\gamma})^{-1} R_{ijt}$ in estimating mean parameters to account for the unobserved responses otherwise similar to Y_{ijt}.

The IPW GEE estimates can be obtained using iterative Fisher scoring algorithm as follow:

$$\tilde{\boldsymbol{\beta}}^{(m+1)} = \tilde{\boldsymbol{\beta}}^{(m)} + \left(\sum_{i=1}^{N} \boldsymbol{D}_i^T(\tilde{\boldsymbol{\beta}}^{(m)}) \Sigma_i(\alpha, \tilde{\boldsymbol{\beta}}^{(m)})^{-1} \Delta_{\beta i}(\hat{\gamma}) \boldsymbol{D}_i(\tilde{\beta}^{(m)})\right)^{-1} \sum_{i=1}^{N} \boldsymbol{D}_i^T(\tilde{\boldsymbol{\beta}}^{(m)}) \Sigma_i(\alpha, \tilde{\boldsymbol{\beta}}^{(m)})^{-1} \Delta_{\beta i}(\hat{\gamma}) \{\mathbf{Z_i} - \boldsymbol{\pi}_i(\tilde{\boldsymbol{\beta}}^{(m)})\} \tag{3.16}$$

Robins et al. (1995)[24] presented the equation to estimate variance of regression estimates $\hat{\boldsymbol{\beta}}$ as follows:

$$var(\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) = \Gamma^{-1}\left(\mathbf{I} - \mathbf{B}\boldsymbol{\Omega}\mathbf{B}^T\right)\Gamma^{-1} \tag{3.17}$$

where $\Gamma = E(\partial U_i(\alpha, \beta, \gamma)/\partial \beta^T)$, $\mathbf{I} = E(U_i(\alpha, \beta, \gamma)U_i(\alpha, \boldsymbol{\beta}, \gamma)^T)$, $\Omega = \left[E(-\partial W_i(\gamma)/\partial \gamma^T))\right]^{-1}$ and $\mathbf{B} = E(\partial U_i(\alpha, \beta, \gamma)/\partial \gamma^T)$. But this variance estimator may not be positive definite. Yi and Cook (2002)[37] provided a positive definitely variance formula derived from (3.17) as follows:

$$var(\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) = \Gamma^{-1}\boldsymbol{\Sigma}[\Gamma^{-1}]^T \tag{3.18}$$

where $\Gamma = E(\partial U_i(\alpha, \beta, \gamma)/\partial \beta^T)$, $\boldsymbol{\Sigma} = E(Q_i(\alpha, \beta, \gamma)Q_i(\alpha, \beta, \gamma)^T)$ and $Q_i(\alpha, \beta, \gamma) = U_i(\alpha, \beta, \gamma) - E(\partial U_i(\alpha, \beta, \gamma)/\partial \gamma^T)\left[E(\partial W_i(\gamma)/\partial \gamma^T))\right]^{-1} W_i(\gamma)$.

3.3.3 ESTIMATING EQUATIONS AND INFERENCE FOR ASSOCIATION PARAMETERS

In section 2.3.2, we proposed to estimate association parameters by minimizing the sum of Euclidean distance between the model based correlation $\boldsymbol{\Sigma}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and the empirical correlation $\mathbf{S_i}(\boldsymbol{\beta}) = (\mathbf{Z_i} - \boldsymbol{\pi}_i(\boldsymbol{\beta}))(\mathbf{Z_i} - \boldsymbol{\pi}_i(\boldsymbol{\beta}))'$. The empirical correlation is expressed in terms of pairwise product of responses minus their expectations. Let $s_{ijtk,j't'k'}$ be the qth elements in $\mathbf{S_i}(\beta)$, where $s_{ijtk,j't'k'} = (z_{ijtk} - \pi_{ijtk}) \times (z_{ij't'k'} - \pi_{ij't'k'})$ and $\sigma_{i,jtk,j't'k'}$ be the qth elements in $\boldsymbol{\Sigma}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})$. The corresponding qth Euclidean distance $(s_{ijtk,j't'k'} - \sigma_{i,jtk,j't'k'})$ should be weighted by the inverse probability of observing a pair of $(z_{ijtk}, z_{ij't'k'})$ or observing a pair of $(Y_{ijt}, Y_{ij't'})$ to account for unobserved pairs or partial observed pairs. Therefore, the inverse probability weighted estimating equation for association parameters is given as:

$$\hat{\boldsymbol{\alpha}} = \text{argmin} \sum_{i=1}^{N} (\mathbf{s}_i(\boldsymbol{\beta}) - \boldsymbol{\sigma}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}))^T \Delta_{\alpha i}(\hat{\gamma})(\mathbf{s}_i(\boldsymbol{\beta}) - \boldsymbol{\sigma}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})) \tag{3.19}$$

This is equivalent to solving the following weighted generalized estimating equation for $\boldsymbol{\alpha}$:

$$\sum_{i=1}^{N} \frac{\partial \boldsymbol{\sigma}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \boldsymbol{\alpha}}^T \Delta_{\alpha i}(\hat{\gamma})\{\mathbf{s}_i(\boldsymbol{\beta}) - \boldsymbol{\sigma}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})\} = 0, \tag{3.20}$$

where $\Delta_{\alpha i}(\hat{\gamma})$ is a diagonal weighting matrix for association parameters. For the qth elements of Euclidean distance $(s_{ijtk,j't'k'} - \sigma_{i,jtk,j't'k'})$, the corresponding qth diagonal element of $\Delta_{\alpha i}(\hat{\gamma})$ equals $I(R_{ijt} = 1, R_{ij't'} = 1)/\bar{\lambda}_{i,jt,j't'}$, where $\bar{\lambda}_{i,jt,j't'}$ is given in section 3.2.2.

3.3.4 ITERATIVE ESTIMATION PROCEDURE

To draw inference, we first solve missing parameters γ in estimating equation (3.9). Then, iterative estimation procedure similar to chapter 2 can be carried out to obtain estimates for mean and association parameters. The whole process is summarized in following manner:
Step 1: Solve for missing parameter $\hat{\gamma}$ from generalized estimating equation (3.9);
Step 2: Obtain an initial estimate of $\boldsymbol{\beta}$ denoted by $\hat{\boldsymbol{\beta}}^{(0)}$ by modeling each outcome separately;
Step 3: Solve for $\hat{\boldsymbol{\alpha}}^{(0)}$ by minimizing equation (3.19) with $\boldsymbol{\beta}$ replaced by $\hat{\boldsymbol{\beta}}^{(0)}$;
Step 4: Solve for $\hat{\boldsymbol{\beta}}^{(1)}$ from generalized estimating equation (3.13) with $\boldsymbol{\alpha}$ replaced by $\hat{\boldsymbol{\alpha}}^{(0)}$.
Step 5: Iterate between Steps 3 and 4 until convergence criteria are fulfilled for both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and the solution is denoted as $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$.

When both the missing model and the response model are correctly specified, the final solution $\hat{\boldsymbol{\beta}}$ obtained from the above iterative procedure is consistent and follows an asymptotic normal distribution with mean $\boldsymbol{\beta}$ and a variance-covariance matrix that can be estimated by equation (3.18) proposed by Yi and Cook (2002)[37]. More specifically,

$$var(\hat{\boldsymbol{\beta}}) = N^{-1}\hat{\Gamma}^{-1}\hat{\Sigma}[\hat{\Gamma}^{-1}]^T \tag{3.21}$$

where $\hat{\Gamma} = N^{-1}\sum_{i=1}^{N} D_i^T(\hat{\beta})\Sigma_i(\hat{\alpha},\hat{\boldsymbol{\beta}})^{-1}\Delta_{\beta i}(\hat{\gamma})D_i(\hat{\beta})$, $\hat{\Sigma} = N^{-1}\sum_{i=1}^{N} Q_i(\hat{\alpha},\hat{\beta},\hat{\gamma})Q_i(\hat{\alpha},\hat{\beta},\hat{\gamma})^T$ and $Q_i(\hat{\alpha},\hat{\beta},\hat{\gamma}) = U_i(\hat{\alpha},\hat{\beta},\hat{\gamma}) - \sum_{i=1}^{N}\partial U_i(\hat{\alpha},\hat{\beta},\hat{\gamma})/\partial\gamma^T[\sum_{i=1}^{N}\partial W_i(\hat{\gamma})/\partial\gamma^T]^{-1}W_i(\hat{\gamma})$ are plug in estimates for Γ, Σ and $Qi(\alpha,\beta,\gamma)$ defined in section 3.3.2.

3.4 SIMULATION STUDY

3.4.1 DESIGN OF SIMULATION STUDY

To examine the performance of IPWGEE in providing unbiased inference for the proposed joint model when data are MAR, we performed a series of simulation studies. For simplicity, we assumed that each subject i has ordinal longitudinal outcome $\boldsymbol{Y_{i1}}$ with three categories and binary longitudinal outcome $\boldsymbol{Y_{i2}}$. We constructed 7 repeated measurements for the ordinal outcome at times $t = 0, 1, 2, 3, 4, 5, 6$ as $\boldsymbol{Y_{i1}}$=$(Y_{i10}, Y_{i11}, Y_{i12}, Y_{i13}, Y_{i14}, Y_{i15}, Y_{i16})'$ and 4 repeated measurements for the binary outcome at times $t' = 0, 2, 4, 6$ as $\boldsymbol{Y_{i2}}$=$(Y_{i20}, Y_{i22}, Y_{i24}, Y_{i26})'$. Accordingly, each subject has a 11×1 dimensional response vector $\boldsymbol{Y_i = (Y'_{i1}, Y'_{i2})'}$, which is generated based on the procedure outlined in section 2.4. For each subject, a time covariate t was specified as above and a baseline covariate X_i was generated from a Bernoulli(0.5) distribution. To specify the true marginal probability model for ordinal outcome $\boldsymbol{Y_{i1}}$ and the true marginal probability model for binary outcome $\boldsymbol{Y_{i2}}$, cumulative logistic regression and logistic regression is used as follows:

Ordinal Outcome:

$$\gamma_{ijtk} = Pr(Y_{ijt} \leq k) = \frac{exp(a_{1k} + \beta_{1t} \times t + \beta_{1x} \times X_i)}{1 + exp(a_{1k} + \beta_{1t} \times t + \beta_{1x} \times X_i)}, \tag{3.22}$$

$$\pi_{ijtk} = Pr(Y_{ijt} = k) = \gamma_{ijtk} - \gamma_{ijtk-1}. \tag{3.23}$$

Binary Outcome:

$$\gamma_{ijt0} = Pr(Y_{ijt} = 0) = \frac{exp(a_{20} + \beta_{2t} \times t + \beta_{2x} \times X_i)}{1 + exp(a_{20} + \beta_{2t} \times t + \beta_{2x} \times X_i)}, \tag{3.24}$$

$$\pi_{ijt0} = Pr(Y_{ijt} = 0) = \gamma_{ijt0}. \tag{3.25}$$

We generated data with two different correlation structures $R(\alpha)$=cov($\boldsymbol{Y_i}$,$\boldsymbol{Y_i}$) as follows: (1) AR type correlation with $\alpha_{11} = 0.8, \alpha_{22} = 0.7, \alpha_{12} = 0.4, \alpha_{212} = 0.3$:

$$R(\alpha) = \begin{pmatrix} 1 & 0.8 & 0.64 & 0.512 & 0.410 & 0.328 & 0.262 & 0.300 & 0.048 & 0.008 & 0.001 \\ 0.8 & 1 & 0.8 & 0.64 & 0.512 & 0.410 & 0.328 & 0.120 & 0.120 & 0.019 & 0.003 \\ 0.64 & 0.8 & 1 & 0.8 & 0.64 & 0.512 & 0.410 & 0.048 & 0.300 & 0.048 & 0.008 \\ 0.512 & 0.64 & 0.8 & 1 & 0.8 & 0.64 & 0.512 & 0.019 & 0.120 & 0.120 & 0.019 \\ 0.410 & 0.512 & 0.64 & 0.8 & 1 & 0.8 & 0.640 & 0.008 & 0.048 & 0.300 & 0.048 \\ 0.328 & 0.410 & 0.512 & 0.64 & 0.8 & 1 & 0.8 & 0.003 & 0.019 & 0.120 & 0.120 \\ 0.262 & 0.328 & 0.410 & 0.512 & 0.64 & 0.8 & 1 & 0.001 & 0.008 & 0.048 & 0.300 \\ 0.300 & 0.120 & 0.048 & 0.019 & 0.008 & 0.003 & 0.001 & 1 & 0.49 & 0.24 & 0.118 \\ 0.048 & 0.120 & 0.300 & 0.120 & 0.048 & 0.019 & 0.008 & 0.49 & 1 & 0.49 & 0.24 \\ 0.008 & 0.019 & 0.048 & 0.120 & 0.300 & 0.120 & 0.048 & 0.24 & 0.49 & 1 & 0.49 \\ 0.001 & 0.003 & 0.008 & 0.019 & 0.048 & 0.120 & 0.300 & 0.118 & 0.24 & 0.49 & 1 \end{pmatrix}, \tag{3.26}$$

(2) Unstructured correlation structure:

$$R(\alpha) = \begin{pmatrix} 1 & 0.5 & 0.4 & 0 & 0 & 0 & 0 & 0.4 & 0.3 & 0 & 0 \\ 0.5 & 1 & 0.5 & 0.4 & 0 & 0 & 0 & 0.35 & 0.35 & 0 & 0 \\ 0.4 & 0.5 & 1 & 0.5 & 0.4 & 0 & 0 & 0.3 & 0.4 & 0.3 & 0 \\ 0 & 0.4 & 0.5 & 1 & 0.5 & 0.4 & 0 & 0 & 0.35 & 0.35 & 0 \\ 0 & 0 & 0.4 & 0.5 & 1 & 0.5 & 0.4 & 0 & 0.3 & 0.4 & 0.3 \\ 0 & 0 & 0 & 0.4 & 0.5 & 1 & 0.5 & 0 & 0 & 0.35 & 0.35 \\ 0 & 0 & 0 & 0 & 0.4 & 0.5 & 1 & 0 & 0 & 0.3 & 0.4 \\ 0.4 & 0.35 & 0.3 & 0 & 0 & 0 & 0 & 1 & 0.4 & 0 & 0 \\ 0.3 & 0.35 & 0.4 & 0.35 & 0.3 & 0 & 0 & 0.4 & 1 & 0.4 & 0 \\ 0 & 0 & 0.3 & 0.35 & 0.4 & 0.35 & 0.3 & 0 & 0.4 & 1 & 0.4 \\ 0 & 0 & 0 & 0 & 0.3 & 0.35 & 0.4 & 0 & 0 & 0.4 & 1 \end{pmatrix}. \tag{3.27}$$

In this simulation study, we only focused on monotone missing pattern. We assumed the first observations of both outcomes are always observed and the missing probability of any response Y_{ijt} only depends on its previous response Y_{ijt-1}. First, let us denote R_{ijt} to be the response indicator for observation Y_{ijt} where $R_{ijt} = 1$ when Y_{ijt} is observed and $R_{ijt} = 0$ when Y_{ijt} is missing. Then the missing mechanism is defined as:

$$logit(R_{ijt} = 1 \mid R_{ijt-1} = 1) = \gamma_j + \gamma \times Y_{ijt-1}. \tag{3.28}$$

We constructed the missing processes of two longitudinal outcomes to be independent, which means the missing of Y_{ijt} does not depend on $Y_{ij't'}$ $(j \neq j')$. When $\gamma = 0$, the missing probability of Y_{ijt} does not depend on previously observed responses, therefore, data are missing complete at random (MCAR). Alternatively, when $\gamma \neq 0$, the missing probability of Y_{ijt} depends on the previously observed responses, therefore, data are missing at random(MAR). Accordingly, One can increase the dependence of missing probability on previous observations, or the level of MAR, by increasing γ value. We examined moderate level of MAR (i.e. $\gamma = 1$) and high level of MAR (i.e. $\gamma = 2.5$). We used γ_j to control the missing percentage, when γ was fixed. Based on the above, the monotone missing patterns of two longitudinal outcomes were generated separately in the following fashion:

For ordinal outcome:

$T = 0 :$ $R_{i10} = 1$ $\forall i$.

$T = 1 :$ p_{i11}=expit$(\gamma_1 + \gamma \times Y_{i10})$, generate R_{i11}~Bernoulli(p_{i11}).

$T = 2 :$ If $R_{i11} = 0$ then $R_{i12} = 0$, otherwise p_{i12}=expit$(\gamma_1+\gamma\times Y_{i11})$, generate R_{i12}~Bernoulli(p_{i12}).

$\vdots$

$T = 6 :$ If $R_{i15} = 0$ then $R_{i16} = 0$, otherwise p_{i16}=expit$(\gamma_1+\gamma\times Y_{i15})$, generate R_{i16}~Bernoulli(p_{i16}).

For binary outcome:

$T = 0 :$ $R_{i20} = 1$ $\forall i$.

$T = 2 :$ p_{i22}=expit$(\gamma_2 + \gamma \times Y_{i20})$, generate R_{i22}~Bernoulli(p_{i22}).

$T = 4 :$ If $R_{i22} = 0$ then $R_{i24} = 0$, otherwise p_{i24}=expit$(\gamma_2+\gamma\times Y_{i22})$, generate R_{i24}~Bernoulli(p_{i24}).

$T = 6 :$ If $R_{i24} = 0$ then $R_{i26} = 0$, otherwise p_{i26}=expit$(\gamma_2+\gamma\times Y_{i24})$, generate R_{i26}~Bernoulli(p_{i26}).

In this simulation study, we compared the performance of 6 different GEE inference procedures for proposed joint model including IPWGEE methods and standard GEE methods:

(1) CC-IND: Joint model fitted with independent correlation structure using complete data.

(2) CC-AR: Joint model fitted with AR type correlation structure using complete data.

(3) AC-IND: Joint model fitted with independence correlation using all available data.

(4) AC-AR: Joint model fitted with AR type correlation using all available data.

(5) IPW-IND: Joint model fitted with independence correlation and inverse probability weighting using all available data.

(6) IPW-AR: Joint model fitted with AR type correlation and inverse probability weighting using all available data.

Inference procedure for standard GEE methods is given in section 2.3. Inference procedure for IPWGEE is given in section 3.3, where missing model needs to be fitted first to construct the inverse probability weights for each observation. Notice that only observations with observed previous outcome should be included in the fitting missing model, in other words, observations at the first time points should not be included.

3.4.2 SIMULATION RESULTS

Table 7 shows the simulation results for data with AR type true correlation structure and 20% missing (i.e. 80% observed responses). In particular, the table provides the true parameter(β), Monte-Carlo average of the estimates ($\hat{\beta}$), Monte-Carlo variance ($var(\hat{\beta})$), Monte-Carlo average of the variance estimates ($\hat{var}(\hat{\beta})$) and the coverage probability (CP%) of 95% confidence intervals. We generated MCAR (i.e. $\gamma = 0$) data and two levels of MAR (i.e. $\gamma = 1$, $\gamma = 2.5$) data and examined 6 different GEE based inference procedures for the proposed joint model including complete case GEE without IPW (i.e. CC-IND, CC-AR), available case GEE without IPW (i.e. AC-IND, AC-AR) and available case GEE with IPW (i.e. IPW-IND, IPW-CC).

When data are MCAR (i.e $\gamma = 0$), naturally, all GEE based estimators for all parameters are unbiased, even using only complete data. The relative bias for β_{1x} and β_{1t} are less than 1% for all fitted joint models. The relative bias for β_{2x} and β_{2t} range from 0.5% to 2% and 1% to 2.6% respectively. The variances of all estimators from complete case GEE are approximately twice as high as the corresponding variance from available case GEE and IPWGEE. The Monte-Carlo average of variance estimates matches the Monte-Carlo variance. The coverage probabilities of 95% CI match the nominal level.

When data are MAR at moderate level (i.e $\gamma = 1$), estimators from complete case GEE are biased. The relative bias for β_{1x}, β_{1t}, and β_{2t} are approximately 8%, 15%, and 12%. The relative bias for β_{2x} is small, and ranges from 2% to 5%. On the other hand, one might expect estimators from available case GEE should also be biased. As Table 7 shows, estimators from available case GEE with independent correlation are biased. However, estimators from available case GEE with correct AR type correlation are unbiased and provide valid coverage probability for 95% confidence intervals, where the relative bias for all parameters range from 0.75% to 3% and the coverage probabilities range from 94.0% to 96.3%. This is discussed in further detail in section 3.5. Finally, estimators from IPWGEE are always unbiased even when fitted with independent correlation structure. The relative bias for all parameters are less than 1%. The variances of complete case GEE estimates are still approximately two to three times higher than the corresponding variance of available case GEE and IPWGEE estimators. Although, the relative bias of complete case GEE estimators are up to 15%, the large variance of complete case GEE generally lead to coverage probabilities that are not too far from 95% nominal level.

When data are MAR at high level (i.e. $\gamma = 2.5$), GEE estimators using only complete data are highly biased. The relative bias for β_{1x}, β_{1t}, β_{2x} and β_{2t} are around 20%, 40%, 13% and 45% respectively. The coverage probabilities are as low as 54.4%. With high MAR level (i.e. $\gamma = 2.5$), both of the available case GEE estimators are biased to a certain extent. However, available case GEE estimators fitted with correct AR type correlation structure are still less biased than those fitted with independent correlation. The relative bias for β_{1x}, β_{1t}, β_{2x} and β_{2t} estimated from AR-AC are 2.6%, 15%, 2% and 1.6% respectively, which are consistently smaller than the corresponding relative bias estimated from AR-IND 9.7%, 42.5%, 4.5% and 5.8%. Finally, both IPWGEE estimators for proposed joint model still provide unbiased estimates and valid coverage probabilities. Furthermore, we observed larger bias with higher level of MAR (i.e. $\gamma = 2.5$) compared to lower level of MAR (i.e. $\gamma = 1$). This is discussed in further detail in section 3.5.

Table 8 shows similar findings for data with unstructured correlation. When data are MCAR, all GEE based estimators are unbiased. When data are MAR, complete case GEE estimators are biased and IPWGEE estimators are unbiased. One important observation is that, although AR type correlation is not the correct correlation in this scenario, standard GEE inference assuming AR type correlation is still less biased than standard GEE inference assuming independence. This demonstrates that although it is difficult to specify a correct correlation, a reasonably close correlation structure can still reduce bias of standard GEE inference due to MAR data.

Table 9 shows the correlation parameter estimates for data with AR type correlation. We observed that second order inverse probability weighting method only improved the estimators for within outcome correlation parameters (i.e. α_{11}, α_{22}). For example, when $\gamma = 0$, $\gamma = 1$ and $\gamma = 2.5$, the relative bias for α_{11} estimated from IPWGEE are 0.3%, 0.4% and 1.5%, which is consistently smaller than the corresponding relative bias of available case GEE estimators 0.4%, 7.2% and 17.5%. However, there is no obvious improvement for estimators for between outcome correlation parameters (i.e. $\alpha_{12}, \alpha_{212}$) and α_{12} is consistently underestimated in Table 9. We suspect that it might require more data points to get reasonable estimates for between correlation parameters.

3.5 DISCUSSION

In chapter 2, we presented a joint model with standard GEE inference for joint analysis of multivariate ordinal longitudinal outcome. The proposed estimator will be biased if data are missing at random. In this chapter, we extend the standard GEE using inverse probability weighting to obtain consistent estimating equations for regression parameters in the presence of missing at random data.

The simulation results showed that joint models with inverse probability weighted estimating equations give consistent estimators for mean parameters when both missing model and response model are correctly specified. Furthermore, simulation also showed that a reasonable correlation structure can reduce bias of standard GEE estimator as well. These observations indicate that both efforts of specifying better missing model and better correlation structure will reduce bias in analysis of missing at random longitudinal data.

When data are missing completely at random, all GEE estimators for the proposed joint model are unbiased and maintains nominal coverage rate for 95% confidence intervals, even with only complete case analysis. Because MCAR data can be viewed as a random sample from the complete dataset, missing data are not expected to introduce bias to inference procedures including GEE. However, the complete case analysis is less efficient compared to other methods because it ignores the subjects with any missing data and consequently, cause a reduction in sample size.

When data are missing at random, complete case GEE estimators are biased. Furthermore, we observe larger bias with higher level of MAR compared to lower level of MAR. Because our construction of the missing model implies that smaller values of γ are closer to MCAR, this result is expected.

One might expect estimators from available case GEE to be biased as well. However, with moderate MAR level, available case GEE estimators with the correct correlation structure are able to provide unbiased estimates and valid coverage probabilities. Even with high MAR level, estimators for available case GEE fitted with correct correlation structure are less biased than those fitted with independent correlation. Furthermore, we observed that even standard GEE fitted with incorrect but reasonably close correlation structure helped to reduce bias. This is because, for correlated longitudinal data, when there are missing responses for a subject, previous observed responses from this subject will provide information about the unobserved response if the correlation structure is correctly specified or is reasonably close. Therefore, assuming reasonable correlation decreases the bias or even provide unbiased estimators with moderate MAR level. This needs to be distinguished from complete case GEE estimators, where assuming correct correlation structure does not reduce bias. Because subjects are independent and knowing about subjects with complete data is unable to provide any information for subjects with missing data. As we expected, with complete case analysis, the magnitude of bias is very similar for both estimators assuming correct correlation structure and independence. One may also argue that, GEE method with correct correlation structure has not only the correct marginal mean model but also the correct correlation structure, which is similar to the likelihood based methods with correct distributional assumption. Therefore, we expect that GEE with reasonable correlation assumption behaves more similarly to the likelihood based methods, which provides consistent estimators when data are MAR. This observation demonstrates the importance of correlation parameters in analysis of missing at random longitudinal data. It offers another way to reduce bias for standard GEE method when data are MAR, in addition to using inverse probability weighting methods.

IPWGEE estimators are unbiased, even when fitted with independent correlation, which is a highly desirable property of IPWGEE estimators. Because even though correct or reasonably close correlation structure reduce bias for standard GEE method using available data as well. In practice, it might be difficult to specify the correct or reasonably close correlation structure for longitudinal outcomes. However, it is important to note that both the efforts of specifying better missing model and better correlation structure will reduce bias in analysis of missing at random longitudinal data. This indicates two possible ways to reduce bias for standard GEE method when data are MAR.

Finally, in terms of efficiency, when all available data are used in the inference procedure, IPWGEE estimators have slightly higher variance compared with standard GEE estimators. Therefore, using IPW methods to correct bias, when data are MAR, leads to slight efficiency loss.

4.0 FUTURE WORK

In this dissertation, we first present a joint model for analysis of multivariate ordinal longitudinal outcome, which assumes ordered outcomes arose from a partitioned latent multivariate normal process. This joint model provides a framework for analyzing multivariate ordered longitudinal data with a general multilevel association structure, covering both between and within outcome correlation. Simulation studies show that the estimators of mean parameters are unbiased and more efficient than those obtained through fitting separate standard GEE for each outcome. The proposed method also yields unbiased estimators for correlation parameters given the correct correlation structure are specified. However, GEE based estimators are biased when missing data are present and the missing mechanism is not missing completely at random (MCAR). In the second part of this dissertation, we extend our joint model to handle missing at random (MAR) data by using inverse probability weighted (IPW) second order estimating equations. Simulation studies show that IPW estimators remain consistent when both the missing model and response model are correctly specified.

4.1 DOUBLE ROBUST GENERALIZED ESTIMATING EQUATIONS

When data are missing at random, a common modification for GEE is inverse probability weighting (IPW) which assumes certain underlying missing mechanism causes some responses to be observed with lower probabilities. In order to obtain consistent estimators, IPWGEE method weighs responses using the inverse probability of being observed. When data are MAR, IPWGEE estimator provides consistent regression estimates when both the response model and missing model are correctly specified.

Robins et al. (1994) [23] and Rotnitzky et al. (1998)[26] proposed an augmented inverse probability weighted estimator (AIPW), which was proved to be doubly robust by Scharfstein et al. (1999) [29]. Unlike IPWGEE method, doubly robust estimator remains consistent when either the missing model and response model for the complete data is correctly specified.

Later, Bang and Robins (2005) [2] demonstrated the application of such doubly robust estimator in longitudinal data with monotone missing. Accordingly, one possible future work following this dissertation may include the application of recently developed doubly robust methods to analyze multivariate ordinal longitudinal outcome that is missing at random, which could also be extended to data with intermittent or non-monotone missing pattern. As stated by Bang and Robins (2005)[2], one of the most significant advantages of doubly robust estimator is that when either missing model or mean model is almost correct, the bias of doubly robust estimators will be small. Therefore, compared to IPWGEE and imputation method, doubly robust estimators give analysis one more chance to get nearly correct inference.

APPENDIX

TABLES

Table 1: Performance of different methods with extended exchangeable and AR-type correlations. Results are from M=500 datasets with n=100 subjects.

True Correlation	Parameters	$\boldsymbol{\beta}$	sep-GLM[a] $\hat{\boldsymbol{\beta}}$	sep-GLM[a] var($\hat{\boldsymbol{\beta}}$)	sep-GLM[a] $\hat{var}(\hat{\boldsymbol{\beta}})$	sep-GLM[a] CP%	sep-GEE[b] $\hat{\boldsymbol{\beta}}$	sep-GEE[b] var($\hat{\boldsymbol{\beta}}$)	sep-GEE[b] $\hat{var}(\hat{\boldsymbol{\beta}})$	sep-GEE[b] CP%	Joint GEE Independence[c] $\hat{\boldsymbol{\beta}}$	Joint GEE Independence[c] var($\hat{\boldsymbol{\beta}}$)	Joint GEE Independence[c] $\hat{var}(\hat{\boldsymbol{\beta}})$	Joint GEE Independence[c] CP%	Joint GEE[d] $\hat{\boldsymbol{\beta}}$	Joint GEE[d] var($\hat{\boldsymbol{\beta}}$)	Joint GEE[d] $\hat{var}(\hat{\boldsymbol{\beta}})$	Joint GEE[d] CP%	RE
1)	β_{1x}	0.3	0.302	0.140	0.033	67.2%	0.302	0.140	0.134	95.8%	0.303	0.140	0.134	95.8%	0.304	0.132	0.130	95.6%	1.07
Exchangeable	β_{1t}	0.2	0.201	0.001	0.004	100%	0.201	0.001	0.001	94.2%	0.201	0.001	0.001	94.4%	0.201	0.001	0.001	95.2%	1.06
$(\alpha_{11}, \alpha_{22}, \alpha_{12})$	β_{2x}	0.4	0.400	0.213	0.112	86.0%	0.410	0.197	0.196	95.8%	0.402	0.214	0.206	94.4%	0.404	0.160	0.168	96.2%	1.23
$= (0.9, 0.9, 0.9)$	β_{2t}	0.5	0.511	0.011	0.013	97.8%	0.511	0.011	0.010	93.6%	0.511	0.011	0.009	94.2%	0.509	0.010	0.009	94.1%	1.10
2)	β_{1x}	0.3	0.310	0.131	0.033	68.6%	0.310	0.131	0.118	93.2%	0.310	0.131	0.118	93.2%	0.309	0.129	0.114	94.6%	1.02
Exchangeable	β_{1t}	0.2	0.201	0.002	0.004	99.8%	0.201	0.002	0.002	94.2%	0.201	0.002	0.002	94.2%	0.200	0.002	0.002	93.8%	1.03
$(\alpha_{11}, \alpha_{22}, \alpha_{12})$	β_{2x}	0.4	0.411	0.187	0.111	87.0%	0.410	0.175	0.173	95.6%	0.411	0.187	0.178	94.0%	0.408	0.164	0.169	95.8%	1.07
$= (0.8, 0.7, 0.4)$	β_{2t}	0.5	0.510	0.010	0.013	97.4%	0.510	0.010	0.010	93.4%	0.510	0.010	0.010	93.2%	0.509	0.010	0.010	93.6%	1.01
3)	β_{1x}	0.3	0.313	0.034	0.033	93.4%	0.313	0.034	0.033	93.2%	0.313	0.034	0.033	93.2%	0.313	0.034	0.033	93.6%	1.00
Exchangeable	β_{1t}	0.2	0.201	0.004	0.004	96.8%	0.201	0.004	0.004	96.2%	0.201	0.004	0.004	96.2%	0.201	0.004	0.004	96.2%	1.00
$(\alpha_{11}, \alpha_{22}, \alpha_{12})$	β_{2x}	0.4	0.403	0.121	0.109	93.6%	0.403	0.121	0.110	94.0%	0.403	0.121	0.109	93.8%	0.402	0.121	0.109	93.8%	1.00
$= (0, 0, 0)$	β_{2t}	0.5	0.513	0.014	0.013	95.6%	0.513	0.014	0.013	94.8%	0.513	0.014	0.013	95.0%	0.513	0.014	0.013	94.8%	1.00
4)	β_{1x}	0.3	0.303	0.127	0.033	68.4%	0.303	0.127	0.122	95.6%	0.304	0.128	0.122	95.6%	0.306	0.127	0.117	95.0%	1.02
AR-type	β_{1t}	0.2	0.202	0.003	0.004	99.2%	0.202	0.003	0.003	95.8%	0.202	0.003	0.003	95.8%	0.202	0.002	0.003	96.1%	1.01
$(\alpha_{11}, \alpha_{22}, \alpha_{12}, \alpha_{212})$	β_{2x}	0.4	0.410	0.202	0.111	86.6%	0.414	0.193	0.183	96.0%	0.411	0.202	0.192	95.6%	0.395	0.172	0.179	96.1%	1.13
$= (0.9, 0.9, 0.9, 0.9)$	β_{2t}	0.5	0.508	0.011	0.013	98.0%	0.508	0.011	0.010	94.4%	0.508	0.011	0.010	94.2%	0.507	0.010	0.010	93.4%	1.02
5)	β_{1x}	0.3	0.317	0.113	0.033	73.2%	0.317	0.112	0.101	94.4%	0.317	0.113	0.101	94.4%	0.316	0.109	0.096	94.8%	1.03
AR-type	β_{1t}	0.2	0.203	0.004	0.004	96.6%	0.203	0.004	0.004	95.4%	0.203	0.004	0.004	95.4%	0.204	0.004	0.004	95.0%	1.02
$(\alpha_{11}, \alpha_{22}, \alpha_{12}, \alpha_{212})$	β_{2x}	0.4	0.409	0.163	0.110	89.4%	0.410	0.159	0.148	94.4%	0.409	0.163	0.148	93.2%	0.407	0.155	0.144	93.8%	1.03
$= (0.8, 0.7, 0.4, 0.3)$	β_{2t}	0.5	0.510	0.013	0.013	94.8%	0.510	0.013	0.012	93.4%	0.510	0.013	0.012	93.4%	0.509	0.013	0.012	94.0%	1.00
6)	β_{1x}	0.3	0.313	0.034	0.033	93.4%	0.313	0.034	0.033	93.2%	0.313	0.035	0.034	93.6%	0.313	0.034	0.033	93.4%	1.00
AR-type	β_{1t}	0.2	0.201	0.004	0.004	96.8%	0.201	0.004	0.004	96.2%	0.202	0.004	0.004	95.4%	0.201	0.004	0.004	96.4%	1.00
$(\alpha_{11}, \alpha_{22}, \alpha_{12}, \alpha_{212})$	β_{1x}	0.4	0.404	0.121	0.109	93.6%	0.403	0.121	0.110	94.0%	0.403	0.124	0.114	94.4%	0.404	0.120	0.109	94.4%	1.00
$= (0, 0, 0, 0)$	β_{1t}	0.5	0.513	0.014	0.013	95.6%	0.513	0.014	0.013	94.8%	0.514	0.014	0.013	95.2%	0.513	0.014	0.013	94.8%	1.00

sep-GLM[a]: separately fit cumulative logistic regression for ordinal outcome and logistic regression for binary outcome using maximum likelihood; sep-GEE[b]: separately fit the ordinal outcome and the binary outcome using GEE; Joint GEE Independence[c]: GEE applied to the joint model with independent correlation structure; Joint GEE[d]: GEE applied to the joint model where correlation parameter $\boldsymbol{\alpha}$ is estimated along with regression parameter $\boldsymbol{\beta}$.

Table 2: Performance of different methods with extended exchangeable and AR-type correlation. Results are from M=1000 datasets with n=50 subjects.

			sep-GLM[a]				sep-GEE[b]				Joint GEE Independence[c]				Joint GEE[d]				
True Correlation	Parameters	$\boldsymbol{\beta}$	$\hat{\boldsymbol{\beta}}$	var($\hat{\boldsymbol{\beta}}$)	$v\hat{a}r(\hat{\boldsymbol{\beta}})$	CP%	$\hat{\boldsymbol{\beta}}$	var($\hat{\boldsymbol{\beta}}$)	$v\hat{a}r(\hat{\boldsymbol{\beta}})$	CP%	$\hat{\boldsymbol{\beta}}$	var($\hat{\boldsymbol{\beta}}$)	$v\hat{a}r(\hat{\boldsymbol{\beta}})$	CP%	$\hat{\boldsymbol{\beta}}$	var($\hat{\boldsymbol{\beta}}$)	$v\hat{a}r(\hat{\boldsymbol{\beta}})$	CP%	RE
1)	β_{1x}	0.3	0.296	0.301	0.069	65.8%	0.296	0.301	0.273	94.0%	0.297	0.303	0.272	94.0%	0.302	0.299	0.265	94.0%	1.00
Exchangeable	β_{1t}	0.2	0.207	0.003	0.008	99.9%	0.207	0.003	0.003	93.7%	0.207	0.003	0.003	93.8%	0.207	0.003	0.003	94.9%	1.04
$(\alpha_{11}, \alpha_{22}, \alpha_{12})$	β_{2x}	0.4	0.423	0.490	0.242	84.5%	0.431	0.488	0.429	95.0%	0.426	0.492	0.427	94.1%	0.418	0.393	0.349	94.6%	1.24
$= (0.9, 0.9, 0.9)$	β_{2t}	0.5	0.534	0.022	0.029	98.8%	0.534	0.022	0.021	93.2%	0.534	0.022	0.020	93.1%	0.531	0.020	0.019	94.4%	1.09
2)	β_{1x}	0.3	0.284	0.259	0.068	69.4%	0.284	0.259	0.237	94.4%	0.287	0.259	0.237	94.4%	0.285	0.254	0.232	94.6%	1.02
Exchangeable	β_{1t}	0.2	0.208	0.004	0.008	99.8%	0.208	0.004	0.004	93.8%	0.208	0.004	0.004	93.9%	0.208	0.004	0.004	93.3%	1.02
$(\alpha_{11}, \alpha_{22}, \alpha_{12})$	β_{2x}	0.4	0.435	0.399	0.236	87.3%	0.441	0.399	0.367	94.4%	0.440	0.401	0.370	92.9%	0.431	0.365	0.353	95.2%	1.10
$= (0.8, 0.7, 0.4)$	β_{2t}	0.5	0.529	0.023	0.028	98.7%	0.530	0.023	0.021	93.3%	0.529	0.023	0.020	93.0%	0.527	0.022	0.021	93.7%	1.03
3)	β_{1x}	0.3	0.289	0.076	0.067	92.3%	0.289	0.076	0.065	92.0%	0.289	0.076	0.065	92.0%	0.289	0.076	0.065	91.9%	1.00
Exchangeable	β_{1t}	0.2	0.204	0.008	0.008	95.9%	0.204	0.008	0.008	95.4%	0.204	0.008	0.008	95.4%	0.204	0.008	0.008	95.4%	1.00
$(\alpha_{11}, \alpha_{22}, \alpha_{12})$	β_{2x}	0.4	0.434	0.235	0.230	94.9%	0.436	0.238	0.231	94.9%	0.438	0.236	0.227	94.9%	0.435	0.236	0.227	94.7%	1.01
$= (0, 0, 0)$	β_{2t}	0.5	0.522	0.031	0.028	96.0%	0.522	0.032	0.028	95.4%	0.521	0.031	0.027	95.2%	0.522	0.031	0.027	95.3%	1.00
4)	β_{1x}	0.3	0.290	0.268	0.069	68.8%	0.291	0.268	0.246	94.6%	0.293	0.268	0.246	94.6%	0.292	0.269	0.237	93.3%	1.00
AR-type	β_{1t}	0.2	0.205	0.006	0.008	97.3%	0.205	0.006	0.006	93.8%	0.205	0.006	0.006	93.9%	0.205	0.006	0.006	94.1%	1.03
$(\alpha_{11}, \alpha_{22}, \alpha_{12}, \alpha_{212})$	β_{2x}	0.4	0.428	0.452	0.239	86.3%	0.431	0.436	0.385	93.8%	0.430	0.453	0.395	93.6%	0.390	0.374	0.334	94.4%	1.17
$= (0.9, 0.9, 0.9, 0.9)$	β_{2t}	0.5	0.530	0.026	0.029	97.9%	0.530	0.025	0.022	93.0%	0.531	0.026	0.021	91.4%	0.527	0.023	0.020	93.0%	1.10
5)	β_{1x}	0.3	0.278	0.229	0.068	73.4%	0.278	0.229	0.201	93.8%	0.279	0.231	0.201	93.7%	0.277	0.228	0.193	93.0%	1.01
AR-type	β_{1t}	0.2	0.207	0.008	0.008	95.7%	0.207	0.008	0.008	94.4%	0.207	0.008	0.008	94.3%	0.207	0.008	0.007	94.2%	1.03
$(\alpha_{11}, \alpha_{22}, \alpha_{12}, \alpha_{212})$	β_{2x}	0.4	0.432	0.314	0.234	92.2%	0.437	0.310	0.310	95.4%	0.436	0.315	0.308	95.6%	0.430	0.299	0.299	95.5%	1.04
$= (0.8, 0.7, 0.4, 0.3)$	β_{2t}	0.5	0.525	0.029	0.028	96.6%	0.524	0.028	0.026	94.2%	0.525	0.029	0.025	92.5%	0.521	0.028	0.025	94.0%	1.01
6)	β_{1x}	0.3	0.289	0.076	0.067	92.3%	0.289	0.076	0.065	92.0%	0.289	0.076	0.065	92.0%	0.289	0.076	0.065	92.2%	1.00
AR-type	β_{1t}	0.2	0.204	0.008	0.008	95.9%	0.204	0.008	0.008	95.4%	0.204	0.008	0.008	95.4%	0.204	0.008	0.008	95.2%	1.00
$(\alpha_{11}, \alpha_{22}, \alpha_{12}, \alpha_{212})$	β_{2x}	0.4	0.434	0.235	0.230	94.9%	0.435	0.235	0.231	95.2%	0.438	0.236	0.227	94.9%	0.435	0.235	0.225	95.1%	1.00
$= (0, 0, 0, 0)$	β_{2t}	0.5	0.522	0.031	0.028	96.0%	0.522	0.031	0.028	95.3%	0.521	0.031	0.027	95.2%	0.521	0.031	0.027	95.4%	1.00

sep-GLM[a]: separately fit cumulative logistic regression for ordinal outcome and logistic regression for binary outcome using maximum likelihood; sep-GEE[b]: separately fit the ordinal outcome and the binary outcome using GEE; Joint GEE Independence[c]: GEE applied to the joint model with independent correlation structure; Joint GEE[d]: GEE applied to the joint model where correlation parameter $\boldsymbol{\alpha}$ is estimated along with regression parameter $\boldsymbol{\beta}$.

Table 3: Estimates and Monte-Carlo variances of correlation parameters in Tables 1 and 2 estimated from the joint GEE. A) Table 1: M=500 datasets with n=100 subjects; B) Table 2: M=1000 datasets with n=50 subjects.

Correlation Type	Correlation Parameters	$\boldsymbol{\alpha}$	A $\hat{\boldsymbol{\alpha}}$	A $var(\hat{\boldsymbol{\alpha}})$	B $\hat{\boldsymbol{\alpha}}$	B $var(\hat{\boldsymbol{\alpha}})$
Exchangeable	α_{11}	0.9	0.908	0.002	0.892	0.004
	α_{22}	0.9	0.917	0.006	0.903	0.012
	α_{12}	0.9	0.902	0.001	0.899	0.003
	α_{11}	0.8	0.795	0.003	0.778	0.006
	α_{22}	0.7	0.693	0.014	0.683	0.030
	α_{12}	0.4	0.406	0.012	0.405	0.025
	α_{11}	0	0.000	0.003	-0.010	0.005
	α_{22}	0	-0.009	0.024	-0.029	0.051
	α_{12}	0	0.002	0.003	0.005	0.007
AR-type	α_{11}	0.9	0.899	0.001	0.889	0.002
	α_{22}	0.9	0.898	0.023	0.886	0.024
	α_{12}	0.9	0.900	0.001	0.894	0.003
	α_{212}	0.9	0.901	0.002	0.903	0.005
	α_{11}	0.8	0.796	0.003	0.780	0.005
	α_{22}	0.7	0.681	0.016	0.650	0.047
	α_{12}	0.4	0.381	0.127	0.366	0.174
	α_{212}	0.3	0.305	0.018	0.305	0.037
	α_{11}	0	-0.003	0.007	-0.011	0.014
	α_{22}	0	0.017	0.003	0.061	0.026
	α_{12}	0	-0.007	0.030	0.016	0.112
	α_{212}	0	-0.008	0.018	0.003	0.035

Table 4: Performance of the joint model when correlation structure is misspecified. Results are from M=500 datasets with n=100 subjects.

			Joint GEE Independence[a]				Joint GEE Exchangeable[b]				Joint GEE AR[c]			
True Correlation	Parameters	$\boldsymbol{\beta}$	$\hat{\beta}$	$\mathrm{var}(\hat{\beta})$	$v\hat{a}r(\hat{\boldsymbol{\beta}})$	CP%	$\hat{\beta}$	$\mathrm{var}(\hat{\beta})$	$v\hat{a}r(\hat{\boldsymbol{\beta}})$	CP%	$\hat{\beta}$	$\mathrm{var}(\hat{\beta})$	$v\hat{a}r(\hat{\boldsymbol{\beta}})$	CP%
	β_{1x}	0.3	0.310	0.131	0.118	93.2%	0.309	0.129	0.114	94.6%	0.308	0.132	0.117	94.8%
Exchangeable	β_{1t}	0.2	0.201	0.002	0.002	94.2%	0.200	0.002	0.002	93.8%	0.200	0.002	0.002	93.6%
$(\alpha_{11}, \alpha_{22}, \alpha_{12})$	β_{2x}	0.4	0.411	0.187	0.178	94.0%	0.408	0.164	0.169	95.8%	0.409	0.164	0.170	96.0%
$= (0.8, 0.7, 0.4)$	β_{2t}	0.5	0.510	0.010	0.010	93.2%	0.509	0.010	0.010	93.6%	0.509	0.010	0.010	94.0%
	β_{1x}	0.3	0.317	0.113	0.101	94.4%	0.317	0.111	0.099	94.2%	0.316	0.109	0.096	94.8%
AR-type	β_{1t}	0.2	0.203	0.004	0.004	95.4%	0.203	0.004	0.004	94.6%	0.204	0.004	0.004	95.0%
$(\alpha_{11}, \alpha_{22}, \alpha_{12}, \alpha_{212})$	β_{2x}	0.4	0.409	0.163	0.148	93.2%	0.406	0.156	0.147	94.8%	0.407	0.155	0.144	93.8%
$= (0.8, 0.7, 0.4, 0.3)$	β_{2t}	0.5	0.510	0.013	0.012	93.4%	0.510	0.013	0.012	93.6%	0.509	0.013	0.012	94.0%

Joint GEE Independence[a]: GEE applied to the joint model with independent correlation structure; Joint GEE Exchangeable[b]: GEE applied to the joint model with extended exchangeable correlation structure where correlation parameter $\boldsymbol{\alpha}$ is estimated along with regression parameter $\boldsymbol{\beta}$; Joint GEE AR[c]: GEE applied to the joint model with extended AR-type correlation structure where correlation parameter $\boldsymbol{\alpha}$ is estimated along with regression parameter $\boldsymbol{\beta}$;

Table 5: Number of observed patients at each week.

Week	Peginterferon(%)	Ribavirin(%)
0	388(100%)	389(100%)
1	388(100%)	389(100%)
2	388(100%)	389(100%)
3	388(100%)	389(100%)
4	379(97.6%)	389(100%)
5	376(96.9%)	380(97.7%)
6	375(96.6%)	377(96.9%)
7	375(96.6%)	375(96.4%)
8	373(96.1%)	375(96.4%)
9	372(95.9%)	373(95.9%)
10	372(95.9%)	370(95.1%)
11	371(95.6%)	367(94.3%)
12	365(94.1%)	367(94.3%)

Table 6: Regression estimates (standard errors) and p value for Virahep-C study, using sep-GEE and Joint GEE assuming an extended AR-type correlation.

	Covariates	sep-GEE[a]		Joint GEE[b]	
		Est(SE)	p value	Est(SE)	p value
Peginterferon (Binary)	Intercept	-3.899(0.903)	$< .0001$	-3.600(0.922)	$< .0001$
	CA race	-0.660(0.298)	0.0266	-0.569(0.294)	0.0526
	days	0.006(0.003)	0.0326	0.004(0.003)	0.1616
	CA race*days	-0.001(0.005)	0.8918	-0.001(0.005)	0.8931
	sex	-0.112(0.284)	0.6924	-0.064(0.282)	0.8194
	vloadblg	0.087(0.145)	0.5474	0.049(0.151)	0.7425
	employ	0.882(0.273)	0.0012	0.812(0.267)	0.0023
Ribavirin (Ordinal)	Intercept1	-3.846(0.746)	$< .0001$	-3.840(0.713)	$< .0001$
	Intercept2	-2.521(0.749)	0.0008	-2.453(0.706)	0.0005
	CA race	-0.559(0.205)	0.0064	-0.322(0.210)	0.1241
	days	0.012(0.002)	$< .0001$	0.014(0.002)	$< .0001$
	CA race*days	-0.002(0.003)	0.5094	-0.003(0.003)	0.1944
	sex	0.023(0.190)	0.9016	0.123(0.189)	0.5159
	vloadblg	0.068(0.115)	0.5560	0.022(0.109)	0.8371
	employ	-0.055(0.186)	0.7666	-0.222(0.158)	0.1600

sep-GEE[a]: separately analyze the ordinal outcome and the binary outcome using GEE; Joint GEE[b]: GEE applied to the joint model where correlation parameter $\boldsymbol{\alpha}$ is estimated along with regression parameter $\boldsymbol{\beta}$.

Table 7: Joint models fitted with different GEE based methods for data with AR type correlation. Results are from 800 datasets with 200 subjects and 20% missing data.

			AR type Correlation											
			$R \approx 80\%, \gamma = 0^a$				$R \approx 80\%, \gamma = 1^a$				$R \approx 80\%, \gamma = 2.5^a$			
Joint Models	Parameters	$\boldsymbol{\beta}$	$\hat{\beta}$	var($\hat{\beta}$)	$\hat{var}(\hat{\boldsymbol{\beta}})$	CP%	$\hat{\beta}$	var($\hat{\beta}$)	$\hat{var}(\hat{\boldsymbol{\beta}})$	CP%	$\hat{\beta}$	var($\hat{\beta}$)	$\hat{var}(\hat{\boldsymbol{\beta}})$	CP%
CC-IND[b]	β_{1x}	0.3	0.303	0.111	0.111	94.9%	0.277	0.116	0.108	94.3%	0.240	0.119	0.117	95.5%
	β_{1t}	0.2	0.202	0.003	0.003	94.5%	0.231	0.003	0.003	93.4%	0.276	0.004	0.004	78.3%
	β_{2x}	0.4	0.398	0.129	0.133	95.6%	0.392	0.138	0.134	94.4%	0.347	0.166	0.165	95.4%
	β_{2t}	0.5	0.513	0.006	0.006	94.3%	0.566	0.007	0.007	88.6%	0.740	0.017	0.016	55.0%
CC-AR[c]	β_{1x}	0.3	0.301	0.104	0.105	95.3%	0.275	0.112	0.102	94.0%	0.238	0.117	0.114	95.5%
	β_{1t}	0.2	0.202	0.003	0.002	94.3%	0.234	0.003	0.003	91.5%	0.281	0.004	0.004	74.9%
	β_{2x}	0.4	0.397	0.127	0.130	95.4%	0.386	0.138	0.129	94.4%	0.344	0.163	0.161	95.4%
	β_{2t}	0.5	0.513	0.006	0.006	94.4%	0.561	0.007	0.007	90.5%	0.729	0.016	0.015	54.4%
AC-IND[d]	β_{1x}	0.3	0.303	0.050	0.049	95.8%	0.288	0.046	0.045	94.9%	0.271	0.039	0.040	95.5%
	β_{1t}	0.2	0.201	0.001	0.001	94.8%	0.158	0.001	0.001	88.5%	0.115	0.001	0.001	38.1%
	β_{2x}	0.4	0.408	0.064	0.067	96.4%	0.397	0.063	0.062	95.0%	0.382	0.056	0.058	94.4%
	β_{2t}	0.5	0.506	0.003	0.003	95.3%	0.493	0.003	0.003	95.7%	0.471	0.003	0.003	90.1%
AC-AR[e]	β_{1x}	0.3	0.298	0.046	0.045	95.4%	0.296	0.047	0.044	94.3%	0.291	0.042	0.042	96.2%
	β_{1t}	0.2	0.202	0.001	0.001	95.0%	0.194	0.001	0.001	94.0%	0.170	0.001	0.001	87.0%
	β_{2x}	0.4	0.408	0.065	0.065	95.8%	0.403	0.063	0.062	95.4%	0.392	0.058	0.060	94.9%
	β_{2t}	0.5	0.506	0.003	0.003	95.3%	0.504	0.003	0.003	96.3%	0.492	0.003	0.003	93.8%
IPW-IND[f]	β_{1x}	0.3	0.303	0.051	0.050	95.5%	0.297	0.054	0.053	94.5%	0.317	0.077	0.074	94.6%
	β_{1t}	0.2	0.202	0.001	0.001	95.3%	0.201	0.001	0.001	96.5%	0.201	0.002	0.002	95.6%
	β_{2x}	0.4	0.407	0.066	0.067	96.1%	0.401	0.068	0.065	94.1%	0.405	0.077	0.077	94.1%
	β_{2t}	0.5	0.506	0.003	0.003	94.8%	0.504	0.003	0.003	96.0%	0.506	0.003	0.003	94.6%
IPW-AR[g]	β_{1x}	0.3	0.298	0.049	0.046	94.9%	0.298	0.054	0.052	94.2%	0.285	0.176	0.084	95.1%
	β_{1t}	0.2	0.202	0.001	0.001	94.9%	0.201	0.001	0.001	93.8%	0.202	0.003	0.002	94.6%
	β_{2x}	0.4	0.407	0.069	0.066	95.1%	0.399	0.069	0.066	95.2%	0.406	0.090	0.095	95.7%
	β_{2t}	0.5	0.505	0.003	0.003	94.5%	0.503	0.003	0.003	95.8%	0.505	0.004	0.004	94.3%

Missing Model[a]: $logit(R_{ijt} = 1 | R_{ijt-1} = 1) = \gamma_j + \gamma \times y_{ijt-1}$; CC-IND[b]:Joint model fitted with independent correlation structure using complete data; CC-AR[c]: Joint model fitted with AR type correlation structure using complete data; AC-IND[d]: Joint model fitted with independence correlation using all available data; AC-AR[e]: Joint model fitted with AR type correlation using all available data; IPW-IND[f]: Joint model fitted with independence correlation and inverse probability weighting using all available data; IPW-AR[g]: Joint model fitted with AR type correlation and inverse probability weighting using all available data;

Table 8: Joint models fitted with different GEE based methods for data with unstructured correlation. Results are from 800 datasets with 200 subjects and 20% missing data.

			Unstructured Correlation											
			$R \approx 80\%, \gamma = 0^a$				$R \approx 80\%, \gamma = 1^a$				$R \approx 80\%, \gamma = 2.5^a$			
Joint Models	Parameters	$\boldsymbol{\beta}$	$\hat{\beta}$	var($\hat{\beta}$)	$\hat{var}(\hat{\boldsymbol{\beta}})$	CP%	$\hat{\beta}$	var($\hat{\beta}$)	$\hat{var}(\hat{\boldsymbol{\beta}})$	CP%	$\hat{\beta}$	var($\hat{\beta}$)	$\hat{var}(\hat{\boldsymbol{\beta}})$	CP%
CC-IND[b]	β_{1x}	0.3	0.303	0.057	0.059	95.8%	0.297	0.064	0.060	95.0%	0.262	0.066	0.061	93.0%
	β_{1t}	0.2	0.202	0.003	0.003	93.9%	0.226	0.003	0.003	93.3%	0.255	0.004	0.004	83.8%
	β_{2x}	0.4	0.397	0.118	0.111	94.3%	0.409	0.136	0.123	93.4%	0.376	0.144	0.140	95.3%
	β_{2t}	0.5	0.509	0.007	0.006	94.1%	0.582	0.009	0.009	88.4%	0.722	0.017	0.015	56.9%
CC-AR[c]	β_{1x}	0.3	0.303	0.055	0.057	95.9%	0.294	0.061	0.058	94.6%	0.263	0.065	0.060	92.8%
	β_{1t}	0.2	0.200	0.003	0.003	94.0%	0.228	0.003	0.003	92.8%	0.259	0.004	0.003	81.3%
	β_{2x}	0.4	0.393	0.118	0.108	93.9%	0.402	0.133	0.117	93.5%	0.370	0.140	0.133	94.8%
	β_{2t}	0.5	0.509	0.007	0.006	94.0%	0.575	0.008	0.008	89.0%	0.695	0.014	0.013	61.8%
AC-IND[d]	β_{1x}	0.3	0.304	0.029	0.028	94.9%	0.298	0.030	0.027	94.1%	0.287	0.027	0.026	94.4%
	β_{1t}	0.2	0.201	0.001	0.001	94.5%	0.188	0.001	0.001	94.7%	0.169	0.001	0.001	88.1%
	β_{2x}	0.4	0.414	0.059	0.057	94.6%	0.409	0.059	0.056	95.3%	0.396	0.052	0.051	95.2%
	β_{2t}	0.5	0.505	0.003	0.003	93.1%	0.499	0.003	0.003	93.0%	0.489	0.003	0.003	92.6%
AC-AR[e]	β_{1x}	0.3	0.302	0.027	0.026	95.1%	0.301	0.028	0.026	93.6%	0.300	0.028	0.027	94.1%
	β_{1t}	0.2	0.201	0.001	0.001	94.0%	0.202	0.001	0.001	95.4%	0.196	0.001	0.001	94.8%
	β_{2x}	0.4	0.401	0.058	0.055	94.1%	0.409	0.059	0.054	95.1%	0.405	0.053	0.052	95.4%
	β_{2t}	0.5	0.506	0.003	0.003	93.1%	0.505	0.003	0.003	93.9%	0.504	0.003	0.003	93.5%
IPW-IND[f]	β_{1x}	0.3	0.304	0.030	0.029	95.4%	0.306	0.032	0.030	93.3%	0.310	0.041	0.042	93.8%
	β_{1t}	0.2	0.201	0.001	0.001	94.0%	0.201	0.001	0.001	95.0%	0.202	0.002	0.002	93.9%
	β_{2x}	0.4	0.413	0.060	0.058	94.6%	0.411	0.062	0.058	94.3%	0.407	0.063	0.066	94.7%
	β_{2t}	0.5	0.505	0.003	0.003	93.0%	0.505	0.003	0.003	94.0%	0.506	0.003	0.003	93.6%
IPW-AR[g]	β_{1x}	0.3	0.303	0.028	0.027	94.6%	0.299	0.030	0.029	93.8%	0.305	0.050	0.046	93.0%
	β_{1t}	0.2	0.201	0.001	0.001	95.1%	0.201	0.001	0.001	94.9%	0.204	0.003	0.002	94.1%
	β_{2x}	0.4	0.410	0.060	0.055	93.9%	0.409	0.063	0.057	93.9%	0.406	0.069	0.071	94.1%
	β_{2t}	0.5	0.505	0.003	0.003	93.1%	0.504	0.003	0.003	94.1%	0.505	0.004	0.004	94.8%

Missing Model[a]: $logit(R_{ijt} = 1 | R_{ijt-1} = 1) = \gamma_j + \gamma \times y_{ijt-1}$; CC-IND[b]:Joint model fitted with independent correlation structure using complete data; CC-AR[c]: Joint model fitted with AR type correlation structure using complete data; AC-IND[d]: Joint model fitted with independence correlation using all available data; AC-AR[e]: Joint model fitted with AR type correlation using all available data; IPW-IND[f]: Joint model fitted with independence correlation and inverse probability weighting using all available data; IPW-AR[g]: Joint model fitted with AR type correlation and inverse probability weighting using all available data.

Table 9: Correlation estimates for data with AR type correlation. Results are from 800 datasets with 200 subjects and 20% missing data.

			AR type Correlation					
			$R \approx 80\%, \gamma = 0^a$		$R \approx 80\%, \gamma = 1^a$		$R \approx 80\%, \gamma = 2.5^a$	
Joint Models	Parameters	$\boldsymbol{\alpha}$	$\hat{\alpha}$	var($\hat{\alpha}$)	$\hat{\alpha}$	var($\hat{\alpha}$)	$\hat{\alpha}$	var($\hat{\alpha}$)
AC-AR[b]	α_{11}	0.8	0.797	0.001	0.742	0.001	0.660	0.002
	α_{22}	0.7	0.665	0.008	0.652	0.006	0.561	0.007
	α_{12}	0.4	0.142	0.005	0.321	0.051	0.294	0.043
	α_{212}	0.3	0.267	0.013	0.294	0.009	0.295	0.008
IPW-AR[c]	α_{11}	0.8	0.798	0.001	0.797	0.002	0.788	0.008
	α_{22}	0.7	0.678	0.007	0.689	0.009	0.668	0.020
	α_{12}	0.4	0.185	0.015	0.189	0.017	0.213	0.043
	α_{212}	0.3	0.306	0.019	0.301	0.020	0.279	0.051

Missing Model[a]: $logit(R_{ijt} = 1|R_{ijt-1} = 1) = \gamma_j + \gamma \times y_{ijt-1}$; AC-AR[b]: Joint model fitted with AR type correlation using all available data; IPW-AR[c]: Joint model fitted with AR type correlation and inverse probability weighting using all available data.

BIBLIOGRAPHY

[1] Rajesh Balkrishnan. Perdictors of medication aherence in the elderly. *Clinical Therapeutics*, 20:764–771, 1998.

[2] Heejung Bang and James M. Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61:962–972, 2005.

[3] Farasat A.S. Bokhari, Frank Heiland, Peter Levine, and G. Thomas Ray. Risk factors for discontinuing drug therapy among children with adhd. *Health Services and Outcomes Research Methodology*, 8:134–158, 2008.

[4] Vincent Carey, Scott L. Zeger, and Peter Diggle. Modelling multivariate binary data with alternating logistic regressions. *Biometrika*, 80:517–526, 1993.

[5] Hari S. Conjeevaram, Michael W. Fried, Lennox J. Jeffers, Norah A. Terrault, Thelma E. Wiley Lucas, Nezam Afdhal, Robert S. Brown, Steven H. Belle, Jay H. Hoofnagle, David E. Kleiner, and Charles D. Howell. Peginterferon and ribavirin treatment in african american patients and caucasian american patients with hepatitis c genotype 1. *Gastroenterology*, 131:470–477, 2006.

[6] Peter Ferenci, Michael W. Fried, Mitchell L. Shiffman, Coleman I. Smith, George Marinos, Fernando L. Gonales Jr., Dieter Hussinger, Moises Diago, Giampero Carosi, Daniel Dhumeaux, Antonio Crax , Monique Chaneac, and K. Rajender Reddy. Predicting sustained virological responses in chronic hepatitis c patients treated with peginterferon alfa-2a (40 kd)/ribavirin. *Journal of Hepatology*, 43:425–433, 2005.

[7] David A. Harville and Robert W. Mee. A mixed model procedure for analyzing ordered categorical data. *Biometrics*, 40:393–408, 1984.

[8] Daniel G. Horvitz and Donovan J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47:663–685, 1952.

[9] Ralph I. Horwitz and Sarah M. Horwitz. Adherence to treatment and health outcomes. *Archives of Internal Medicine*, 153:1863–1868, 1993.

[10] Nan M. Laird. Missing data in longitudinal studies. *Statistics in Medicine*, 7:305–315, 1988.

[11] John C. LaRosa. Poor compliance: the hidden risk factor. *Current Atherosclerosis Reports*, 2:1–4, 2000.

[12] Kung-Yee Liang and Scott L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73:13–22, 1986.

[13] Kung-Yee Liang, Scott L. Zeger, and Bahjat Qaqish. Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society, Series B: Methodological*, 54:3–24, 1992.

[14] Stuart R. Lipsitz, Garrett M. Fitzmaurice, Joseph G. Ibrahim, Debajyoti Sinha, Michael Parzen, and Steven Lipshultz. Joint generalized estimating equations for multivariate longitudinal binary outcomes with missing data: an application to acquired immune deficiency syndrome data. *Journal of the Royal Statistical Society, Series A: Statistics in society*, 172:3–20, 2009.

[15] Michael E. Miller, Charles S. Davis, and J. Richard Landis. The analysis of longitudinal polytomous data: Generalized estimating equations and connections with weighted least squares. *Biometrics*, 49:1033–1044, 1993.

[16] Michela Nose, Corrado Barbui, Richard Gray, and Michele Tansella. Clinical interventions for treatment nonadherence in psychosis: meta-analysis. *The British Journal of Psychiatry*, 183:197–206, 2003.

[17] Lars Osterberg and Terrence Blaschke. Adherence to medication. *The New England Journal of Medicine*, 353:487–497, 2005.

[18] Ross L. Prentice. Correlated binary regression with covariates specific to each binary observation. *Biometrics*, 44:1033–1048, 1988.

[19] Yinsheng Qu, Marion R. Piedmonte, and Sharon V. Medendorp. Latent variable models for clustered ordinal data. *Biometrics*, 51:268–275, 1995.

[20] Yinsheng Qu, George W. Williams, Gerald J. Beck, and Sharon VanderBrug Medendorp. Latent variable models for clustered dichotomous data with multiple subclusters. *Biometrics*, 48:1095–1102, 1992.

[21] M. Raptopoulou, D. Tsantoulas, I. Vafiadi, I. Ketikoglou, E. Paraskevas, T. Vassiliadis, S. Kanatakis, G. Hatzis, L. Sidiropoulos, and E. Akriviadis. The effect of adherence to therapy on sustained response in daily or three times a week interferon alpha-2b plus ribavirin treatment of naive and nonresponder chronic hepatitis c participants. *Journal of Viral Hepatitis*, 12:91–95, 2005.

[22] James M. Robins, Andrea Rotnitzky, and Daniel O. Scharfstein. *Statistical Models for Epidemiology, the Environment, and Clinical Trials*, chapter sensitivity analysis for

selection bias and unmeasured confounding in missing data and causal inference models. In E. Halloran and D. Berry (eds). New York:Springer, 2000.

[23] James M. Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89:846–866, 1994.

[24] James M. Robins, Andrea Rotnitzky, and Lue Ping Zhao. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90:106–121, 1995.

[25] Andrea Rotnitzky and James M. Robins. Analysis of semiparametric regression models with non-ignorable non-response. *Statistics in Medicine*, 16:81–102, 1997.

[26] Andrea Rotnitzky, James M. Robins, and Daniel O. Scharfstein. Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association*, 93:1321–1339, 1998.

[27] Donald B. Rubin. Inference and missing data. *Biometrika*, 63:581–592, 1976.

[28] Donald B. Rubin. Multiple imputations in sample surveys: A phenomenological bayesian approach to nonresponse. *ASA Proceedings of the Section on Survey Research Methods*, pages 20–34, 1978.

[29] Daniel O. Scharfstein, Andrea Rotnitzky, and James M. Robins. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94:1096–1120, 1999.

[30] Mitchell L. Shiffman, Fredy Suter, Bruce R. Bacon, David Nelson, Hugh Harley, Ricard Sola, Stephen D. Shafran, Karl Barange, Amy Lin, Ash Soman, and Stefan Zeuzem. Peginterferon alfa-2a and ribavirin for 16 or 24 weeks in hcv genotype 2 or 3. *The New England Journal of Medicine*, 357:124–134, 2007.

[31] Scott R. Smith, Abdus S. Wahed, Stephanie S. Kelley, Hari S. Conjeevaram, Patricia R. Robuck, and Michael W. Fried. Assessing the validity of self-reported medication adherence in hepatitis c treatment. *The Annals of Pharmacotherapy*, 41:1116–1123, 2007.

[32] Valerie E. Stone, Joseph W. Hogan, Paula Schuman, Anne M. Rompalo, Andrea A. Howard, Christina Korkontzelou, and Dawn K. Smith. Antiretroviral regimen complexity, self-reported adherence, and hiv patients' understanding of their regimens: Survey of women in the her study. *Journal of Acquired Immune Deficiency Syndromes*, 28:124–131, 2001.

[33] Brajendra C. Sutradhar and Milorad Kovacevic. Analysing ordinal longitudinal survey data: Generalised estimating equations approach. *Biometrika*, 87:837–848, 2000.

[34] Edward F. Vonesh and Vernon M. Chinchilli. *Linear and nonlinear models for the analysis of repeated measurements*. CRC Press, 1996.

[35] Bernard Waeber, Gastone Leonetti, Rainer Kolloch, and Gordon T. McInnes. Compliance with aspirin or placebo in the hypertension optimal treatment(hot) study. *Journal of Hypertension*, 17:1041–1045, 1999.

[36] John M. Williamson, KyungMann Kim, and Stuart R. Lipsitz. Analyzing bivariate ordinal data using a global odds ratio. *Journal of the American Statistical Association*, 90:1432–1437, 1995.

[37] Grace Y. Yi and Richard J. Cook. Marginal methods for incomplete longitudinal data arising in clusters. *Journal of the American Statistical Association*, 97:1071–1080, 2002.