# Joint modeling of multiple ordinal adherence outcomes via generalized estimating equations with flexible correlation structure



DOI: 10.1002/sim.7560
Handle: http://hdl.handle.net/1942/26158

**Research Article**

Statistics
in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

# Joint modeling of multiple ordinal adherence outcomes via generalized estimating equations with flexible correlation structure

**Zhen Jiang[a]*, Yimeng Liu[a], Abdus S. Wahed[a] and Geert Molenberghs[b]**

**Adherence to medication is critical in achieving effectiveness of many treatments. Factors influence adherence behavior have been the subject of many clinical studies. Analyzing adherence is complicated because it is often measured on multiple drugs over a period of time, resulting in a multivariate longitudinal outcome. This paper is motivated by the Virahep-C study, where adherence is measured on two drugs as a bivariate ordinal longitudinal outcome. To analyze such outcome, we propose a joint model assuming the multivariate ordinal outcome arose from a partitioned latent multivariate normal process. We also provide a flexible multilevel association structure covering both between and within outcome correlation. In simulation studies, we show that the joint model provides unbiased estimators for regression parameters, which are more efficient than those obtained through fitting separate model for each outcome. The joint method also yields unbiased estimators for the correlation parameters when the correlation structure is correctly specified. Finally, we analyze the Virahep-C adherence data and discuss the findings.** 

**Keywords:** Adherence; Generalized estimating equations; Joint model; Latent variable model; Multivariate ordinal longitudinal data

## 1. INTRODUCTION

### *1.1. Medication Adherence*

Medication adherence is defined as the extent to which patients follow their prescribed treatment regimens. It can be measured using many methods including pill counts, self-reporting, monitoring drug concentration, and electronic pill monitors. Adherence is critical for achieving effectiveness of many medical treatments. Poor adherence often results in lack of treatment effects, worsening of diseases, and increased health care costs.[1–3] Unfortunately, poor adherence is common even in well-monitored clinical trials, especially in treating chronic diseases such as hypertension[4] and psychiatric illness.[5] Analyzing adherence data is complicated because adherence is often measured on multiple drugs over a period of time, resulting in a multivariate longitudinal outcome.

[a] *Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, U.S.A.*
[b] *I-BioStat, Universiteit Hasselt, B-3590 Diepenbeek, Belgium; I-BioStat, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium.*
**Correspondence to: Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, U.S.A. E-mail:zhen.jiang@fda.hhs.gov*

The example motivated this paper is the Viral Resistance to Antiviral Therapy of Chronic Hepatitis C (Virahep-C) study, which assesses the response rates to the peginterferon and ribavirin therapy among African American (AA) and Caucasian American (CA) patients with chronic hepatitis C of genotype 1.[6] The primary objective is to investigate why many patients, more specifically three-fourths of the AA patients, failed to response to the current optimal treatment.[6] Because literature in HCV research has indicated that patients' treatment responses are affected by how closely prescribed treatment regimens are followed,[6–8] one hypothesis is that patients' poor responses could be attributed to poor adherence. Our goal is to identify potential influential factors for medication adherence, and to study the association between adherence to peginterferon and adherence to ribavirin.

In the Virahep-C study, all participants were to receive peginterferon once weekly and ribavirin twice daily. Adherence to peginterferon was dichotomized into adherent or non-adherent. Adherence to ribavirin was categorized as fully adherent, partially adherent, or non-adherent. Therefore, each patient's adherence outcome consists of two longitudinal measurements: one is binary and the other is ordinal. Any patients' adherence status cannot be fully characterized by any one of the two measurements. And these two measurements from the same patient are expected to be correlated and may be affected by similar set of factors. Therefore, rather than fitting separate models, joint modeling of this bivariate longitudinal outcome is preferred. The joint modeling approach may also provides estimates for the correlation between two adherence outcomes, which is of our interest.

## 2. JOINT MODELING

To jointly model the multivariate ordinal longitudinal outcome, one general strategy is to specify its joint density. Verbeke and Davidian,[9] and Molenberghs and Verbeke[10] reviewed several modeling approaches to construct such joint density. The first approach, multivariate marginal models, directly specifies the joint density of the multivariate outcome. Bivariate Plackett distribution,[11] and multivariate distributions within the exponential family[12] have been proposed. This approach gives direct marginal inference, but it requires strong distributional assumptions and would be mathematically challenging for higher dimensions. The second approach, conditional models, specifies a marginal model for one outcome, which the modeling of other outcomes will be conditioned on. Tate,[13] and Little and Schluchter[14] developed different methods following this idea. However, this approach does not provide direct marginal inference. The third approach, shared parameter models, specifies a common random effect in the marginal model for each outcome. Such shared random effect will naturally introduce correlation among multiple outcomes. This approach allows for flexible marginal models and can be easily extended to higher dimensions. However, it also imposes strong restrictions on the association structure, which forces the pairwise correlation between two different outcomes to be equal to the pairwise correlation from one outcome times the correlation from the other outcome.[9] The fourth approach, random-effect models, specifies model specific random effect for each outcome. It enables flexible within and between outcome correlation structure, but causes the dimensions of random effects to increase as the number of outcomes increases. This may not be an issue with fewer outcomes, but it would pose computational issues in fitting models with higher number of outcomes.

The other general strategy, to jointly model the multivariate longitudinal outcome, is to use generalized estimation equation (GEE) based methods, where only the marginal means and the association structure of the joint density are specified. Since the introduction of GEE, numerous improvements have been made for binary and ordinal outcomes. For binary outcomes, correlation[15] and odds ratio[16,17] have been used as the association measure. Qu et al.[18] proposed a latent variable model with a tetrachoric correlation for clustered binary outcomes. For ordinal outcomes, inverse of the Fisher's Z transformation[19] and the global odds ratio[20] have been used as the association measure. Qu et al.[21] also extended their previous work for ordinal outcomes with a polychoric correlation. However, it was limited to univariate ordinal longitudinal outcome.

Joint modeling of the multivariate ordinal longitudinal outcome using GEE based methods is complicated with two

main challenges. The first one is how to link all longitudinal outcomes together and what association measure to use. The second one is how to model the multilevel correlation structure nested in the data (e.g., within and between outcome correlation, subject level correlation, correlation between distinct outcomes at different times, etc.). To analyze this type of data, Lipsitz et al.[22] developed a joint GEE method for the multivariate binary longitudinal outcome. Sutradhar et al.[23] proposed a GEE based method which accounts for the complex between and within outcome correlation as well as the structural correlation from the ordinal nature of the outcomes.

In this paper, we propose a joint modeling approach based on GEE to simultaneously model multiple ordinal longitudinal outcomes. The proposed method models the marginal means and the association structure separately. The marginal probability models can be flexible (e.g. cumulative logit model, probit model). The association structure is assumed to have arisen from a latent multivariate normal process, which provides a flexible framework for multilevel correlation structure. In this paper, we demonstrate the structure of the proposed joint model, illustrate the flexible framework for multilevel correlation structure, and provide the inference procedure in Section 3. In Section 4, we assess the performance of the proposed joint model via simulation studies. In Section 5, we compare the analysis of Virahep-C data based on fitting separate models and the proposed joint model. We conclude the paper with discussions in Section 6.

## 3. THE PROPOSED JOINT MODEL AND INFERENCE

### *3.1. Marginal Probability Models*

Let us denote $Y_{ijt}$ as the $i$th $(i = 1, \ldots, n)$ subject's $j$th $(j = 1, \ldots, J)$ outcome observed at time $t$ $(t = T_{j1}, \ldots, T_{jn_{ij}})$, where $Y_{ijt}$ is a ordinal outcome takes value from $\{0, \ldots, K_j - 1\}$. The response vector $\boldsymbol{Y_i}$ of the $i$th subject is formed as $\boldsymbol{Y_i} = (\boldsymbol{Y_{i1}'}, \ldots, \boldsymbol{Y_{iJ}'})'$, where $\boldsymbol{Y_{ij}}$=$(Y_{ijT_{j1}}, \ldots, Y_{ijT_{jn_{ij}}})'$. Also, let $\mathbf{X_{it}} = (x_{it1}, \ldots, x_{itp})$ be the covariate vector for the $i$th subject at time $t$, $\boldsymbol{\beta}_j = (\beta_{j1}, \beta_{j2}, \ldots, \beta_{jp})'$ be the regression coefficient vector, and $\boldsymbol{a}_j = (a_{j0}, a_{j1}, \ldots, a_{jK_j-2})'$ be the $(K_j - 1)$ dimensional intercept vector for the $j$th outcome.

Let $\gamma_{ijtk} = Pr\left(Y_{ijt} \leq k\right)$ and $\pi_{ijtk} = Pr\left(Y_{ijt} = k\right)$. We assume $\gamma_{ijtk}$ depends on covariates $\mathbf{X_{it}}$ through a cumulative logistic regression model as:

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

$$\begin{aligned}\pi_{ijtk} &= Pr\left(Y_{ijt} = k \mid \mathbf{X_{it}}, \boldsymbol{\beta}_j, \boldsymbol{a}_j\right) = \gamma_{ijtk} - \gamma_{ijtk-1} \\ &= \frac{exp(a_{jk} + \mathbf{X_{it}^T}\boldsymbol{\beta}_j)}{1 + exp(a_{jk} + \mathbf{X_{it}^T}\boldsymbol{\beta}_j)} - \frac{exp(a_{jk-1} + \mathbf{X_{it}^T}\boldsymbol{\beta}_j)}{1 + exp(a_{jk-1} + \mathbf{X_{it}^T}\boldsymbol{\beta}_j)}.\end{aligned} \tag{2}$$

For simplicity, we assume that each outcome has the same set of covariates, although each outcome can have its own set of covariates. We denote $\boldsymbol{\beta} = (\boldsymbol{a}_1', \boldsymbol{\beta}_1', \boldsymbol{a}_2', \boldsymbol{\beta}_2', \ldots, \boldsymbol{a}_J', \boldsymbol{\beta}_J')'$ as the overall regression parameter vector.

### *3.2. Joint Probability Model*

We denote $\gamma_{i,jtk,j't'k'} = Pr\left(Y_{ijt} \leq k, Y_{ij't'} \leq k'\right)$ $(k = 0, \ldots, K_j - 1 \;;\; k' = 0, \ldots, K_j' - 1)$ as the pairwise joint cumulative probability of $Y_{ijt}$ and $Y_{ij't'}$. We assume that $Y_{ijt}$ and $Y_{ij't'}$ originate from a bivariate normal distribution $(u, v) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{ijt,ij't'} \\ \rho_{ijt,ij't'} & 1 \end{pmatrix}\right)$, which is partitioned by threshold values $\Phi^{-1}(\gamma_{ijtk})$ and $\Phi^{-1}(\gamma_{ij't'k'})$. More specifically, $Y_{ijt} \leq k$ when $u \leq \Phi^{-1}(\gamma_{ijtk})$ and $Y_{ij't'} \leq k'$ when $v \leq \Phi^{-1}(\gamma_{ij't'k'})$ . Therefore, $\gamma_{i,jtk,j't'k'}$ and

$\pi_{i,jtk,j't'k'}$ can be expressed as:

$$\gamma_{i,jtk,j't'k'} = Pr(Y_{ijt} \leq k, Y_{ij't'} \leq k') = \int_{-\infty}^{\Phi^{-1}(\gamma_{ijtk})} \int_{-\infty}^{\Phi^{-1}(\gamma_{ij't'k'})} \phi_2(u, v, \rho_{ijt,ij't'})\, \mathrm{d}u\, \mathrm{d}v, \tag{3}$$

$$\pi_{i,jtk,j't'k'} = Pr(Y_{ijt} = k, Y_{ij't'} = k') = \int_{\Phi^{-1}(\gamma_{ijtk-1})}^{\Phi^{-1}(\gamma_{ijtk})} \int_{\Phi^{-1}(\gamma_{ij't'k'-1})}^{\Phi^{-1}(\gamma_{ij't'k'})} \phi_2(u, v, \rho_{ijt,ij't'})\, \mathrm{d}u\, \mathrm{d}v, \tag{4}$$

where $\Phi$ is the cdf of a standard normal distribution, $\phi_2$ is the pdf of a bivariate normal distribution. The correlation $\rho_{ijt,ij't'}$ is also referred to as the polychoric correlation, which is not restricted by marginal probabilities and the number of parameters does not increase as the number of categories of the ordinal outcome increases.[21] We will discuss the structure of $\rho_{ijt,ij't'}$ in section 3.3.

The proposed pairwise joint probability ensures the marginal cumulative probability is still $\Phi\left(\Phi^{-1}(\gamma_{ijtk})\right) = \gamma_{ijtk}$. Only the pairwise joint probability or the second moment is affected by the latent multivariate normal assumption and pairwise correlation $\rho_{ijt,ij't'}$. Therefore, the consistency of marginal regression parameters $\boldsymbol{\beta}$ is not affected even when the latent normal process and/or the correlation structure are misspecified.

This method of modeling joint cumulative probability can be viewed as a multivariate threshold model,[18,21,24] which assumes 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 $\boldsymbol{\varepsilon}_i$, where $\boldsymbol{\varepsilon}_i$ follows a multivariate normal distribution with mean zero and correlation matrix $\mathbf{R}$, and $y_{ijt} = k$ when $\Phi^{-1}(\gamma_{ijtk-1}) < \varepsilon_{ijt} \leq \Phi^{-1}(\gamma_{ijtk})$.

### *3.3. Correlation Structure*

We develop a general framework to model the multilevel association structure, which includes the association within the outcome at different time points, and the association between different outcomes at the same or different time points, or even subject level association. It is motivated by the correlation model proposed by Lipsitz et al.[22] and can be viewed as a natural extension of common univariate correlation structures (e.g. autoregressive(AR), exchangeable, and m-dependent). For example, we construct an extended AR-type correlation as:

$$\rho_{ijt,ij't'} = \alpha_{jj'}^{|t-t'|} \times \alpha_{2jj'}^{I(j \neq j')}, \tag{5}$$

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 at different time points have an AR(1) correlation $\rho_{ijt,ij't'} = \alpha_{jj'}^{|t-t'|}$ ; responses from different outcomes at different time points have a weighted AR(1) correlation $\rho_{ijt,ij't'} = \alpha_{jj'}^{|t-t'|} \times \alpha_{2jj'}$ with weight equals $\alpha_{2jj'}$; and responses from different outcomes at the same time point have correlation $\rho_{ijt,ij't'} = \alpha_{2jj'}$. Similarly, we construct an extended exchangeable correlation and an extended m-dependent correlation as follows:

$$\rho_{ijt,ij't'} = \alpha_{jj'} \times \alpha_{2jj'}^{I(j \neq j')}, \text{ and} \tag{6}$$

$$\rho_{ijt,ij't'} = \alpha_{jj'|t-t'|} \times \alpha_{2jj'}^{I(j \neq j')}. \tag{7}$$

For the extended AR-type and exchangeable correlation in (5) and (6), 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 (7), where the estimation of correlation parameter $\boldsymbol{\alpha}$ may be computationally intensive.

This general framework can be easily extended to model more than two levels of association. For example, when patients at the same hospital are assumed to be clustered, the pairwise correlation structure with three levels of association can be

written as:

$$\rho_{ijt,i'j't'} = \alpha_{jj'}^{|t-t'|} \times \alpha_{2jj'}^{I(j\neq j')} \times \alpha_{3jj'}^{I(i\neq i')}, \tag{8}$$

$$\rho_{ijt,i'j't'} = \alpha_{jj'} \times \alpha_{2jj'}^{I(j\neq j')} \times \alpha_{3jj'}^{I(i\neq i')}, \tag{9}$$

$$\rho_{ijt,i'j't'} = \alpha_{jj'|t-t'|} \times \alpha_{2jj'}^{I(j\neq j')} \times \alpha_{3jj'}^{I(i\neq i')}, \tag{10}$$

where only an extra parameter $\alpha_{3jj'}$ is required. One can construct even more flexible correlation structures by combining different correlation structures at different levels as discussed in Section 6.

In summary, the proposed joint model and correlation structure address both challenges for GEE based methods discussed in Section 2. First, we assume a latent multivariate normal process to link each longitudinal outcome together and use the polychoric correlation as the association measure. Second, we construct a flexible framework to model the multilevel correlation structure.

### *3.4. INFERENCE*

To draw inference, the outcome variable $Y_{ijt}$ with $K_j$ categories is first transformed into a $(K_j - 1)$ dimensional binary vector $Z_{ijt} = (z_{ijt0}, z_{ijt1}, \ldots, z_{ijtK_j-2})$, where

$$z_{ijtk} = \begin{cases} 1 & \text{if } Y_{ijt} = k \\ 0 & \text{if } Y_{ijt} \neq k \quad k = \{0, \ldots, K_j - 2\}, \end{cases} \quad \text{and}$$

$$\pi_{ijtk} = Pr\left(Y_{ijt} = k \mid \mathbf{X_{it}}, \boldsymbol{\beta}_j, \boldsymbol{a}_j\right) \text{ as in (2).}$$

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 $\boldsymbol{\pi}_i$=E($\mathbf{Z_i}$). Our inference procedure is based on this dichotomized outcome variable $\mathbf{Z_i}$.

Had the correlation parameter $\boldsymbol{\alpha}$ been known, the regression parameter $\boldsymbol{\beta}$ can be estimated by solving generalized estimating equations for $\mathbf{Z_i}$ as:

$$\sum_{i=1}^{N} \boldsymbol{D}_i^T(\boldsymbol{\beta})\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\alpha}, \boldsymbol{\beta})\{\mathbf{Z_i} - \boldsymbol{\pi}_i(\boldsymbol{\beta})\} = 0, \tag{11}$$

where $\boldsymbol{\pi}_i(\boldsymbol{\beta})$ is defined in (2), and $\boldsymbol{D}_i(\boldsymbol{\beta}) = \frac{\delta}{\delta\boldsymbol{\beta}}\boldsymbol{\pi}_i(\boldsymbol{\beta})$ is the partial derivative of $\boldsymbol{\pi}_i(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$. For example, the derivative of $\boldsymbol{\pi}_{ijtk}(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}_j$ is given by:

$$\frac{\delta}{\delta\boldsymbol{\beta}_j}\boldsymbol{\pi}_{ijtk}(\boldsymbol{\beta}) = \begin{cases} \mathbf{X_{it}^T}\frac{exp(a_{jk}+\mathbf{X_{it}^T}\boldsymbol{\beta}_j)}{\left(1+exp(a_{jk}+\mathbf{X_{it}^T}\boldsymbol{\beta}_j)\right)^2} & k = 0, \\ \mathbf{X_{it}^T}\left[\frac{exp(a_{jk}+\mathbf{X_{it}^T}\boldsymbol{\beta}_j)}{\left(1+exp(a_{jk}+\mathbf{X_{it}^T}\boldsymbol{\beta}_j)\right)^2} - \frac{exp(a_{jk-1}+\mathbf{X_{it}^T}\boldsymbol{\beta}_j)}{\left(1+exp(a_{jk-1}+\mathbf{X_{it}^T}\boldsymbol{\beta}_j)\right)^2}\right] & k \geq 1. \end{cases} \tag{12}$$

The variance-covariance matrix $\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ of $\mathbf{Z_i}$ can be defined as:

$$\boldsymbol{\Sigma}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \begin{pmatrix} \Sigma_{i11}(\boldsymbol{\alpha}, \boldsymbol{\beta}) & \Sigma_{i12}(\boldsymbol{\alpha}, \boldsymbol{\beta}) & \ldots & \Sigma_{i1J}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \\ \Sigma_{i21}(\boldsymbol{\alpha}, \boldsymbol{\beta}) & \Sigma_{i22}(\boldsymbol{\alpha}, \boldsymbol{\beta}) & \ldots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{iJ1}(\boldsymbol{\alpha}, \boldsymbol{\beta}) & \Sigma_{iJ2}(\boldsymbol{\alpha}, \boldsymbol{\beta}) & \ldots & \Sigma_{iJJ}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \end{pmatrix}, \tag{13}$$

$$\text{where} \quad \Sigma_{ijj'}(\boldsymbol{\alpha},\boldsymbol{\beta}) = \begin{pmatrix} \Sigma_{i,jT_{j1},j'T_{j'1}}(\boldsymbol{\alpha},\boldsymbol{\beta}) & \Sigma_{i,jT_{j1},j'T_{j'2}}(\boldsymbol{\alpha},\boldsymbol{\beta}) & \dots & \Sigma_{i,jT_{j1},j'T_{j'n_{ij'}}}(\boldsymbol{\alpha},\boldsymbol{\beta}) \\ \Sigma_{i,jT_{j2},j'T_{j'1}}(\boldsymbol{\alpha},\boldsymbol{\beta}) & \Sigma_{i,jT_{j2},j'T_{j'2}}(\boldsymbol{\alpha},\boldsymbol{\beta}) & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{i,jT_{jn_{ij}},j'T_{j'1}}(\boldsymbol{\alpha},\boldsymbol{\beta}) & \Sigma_{i,jT_{jn_{ij}},j'T_{j'2}}(\boldsymbol{\alpha},\boldsymbol{\beta}) & \dots & \Sigma_{i,jT_{jn_{ij}},j'T_{n_{ij'}}}(\boldsymbol{\alpha},\boldsymbol{\beta}) \end{pmatrix}. \tag{14}$$

The matrix $\Sigma_{ijt,j't'} = cov(Z_{ijt}, Z_{ij't'})$ is a $(K_j - 1) \times (K'_j - 1)$ dimensional variance-covariance matrix. When $j = j'$ and $t = t'$, $\Sigma_{ijt,j't'}$ represents the structural correlation that arose from the polytomous nature of the ordinal outcome $Y_{ijt}$. When $j \neq j'$ or $t \neq t'$, $\Sigma_{ijt,j't'}$ represents the correlation arose from the latent bivariate normal process for $Y_{ijt}$ and $Y_{ij't'}$.

Let us denote $\sigma_{i,jtk,j't'k'}$ as the $(k,k')$th element of $\Sigma_{ijt,j't'}$. One can show that:

$$\sigma_{i,jtk,j't'k'} = \begin{cases} -\pi_{ijtk}\pi_{ij't'k'} & k \neq k' \\ \pi_{ijtk}(1 - \pi_{ij't'k'}) & k = k' \end{cases} \qquad j = j' \quad \text{and} \quad t = t', \tag{15}$$

$$\sigma_{i,jtk,j't'k'} = \pi_{i,jtk,j't'k'} - \pi_{ijtk}\pi_{ij't'k'} \qquad j \neq j' \quad \text{or} \quad t \neq t', \tag{16}$$

where $\pi_{ijtk}$ and $\pi_{ij't'k'}$ are the marginal probabilities given by (2) and $\pi_{i,jtk,j't'k'}$ is the pairwise joint probability given by (4). Thus, $\boldsymbol{\Sigma}_i$ depends on the correlation parameter $\boldsymbol{\alpha}$ through the assumed latent random normal process.

Now, we can solve (11) using an iterative equation:

$$\tilde{\boldsymbol{\beta}}^{(m+1)} = \tilde{\boldsymbol{\beta}}^{(m)} + \left( \sum_{i=1}^{N} \boldsymbol{D}_i^T(\tilde{\boldsymbol{\beta}}^{(m)}) \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\alpha}, \tilde{\boldsymbol{\beta}}^{(m)}) \boldsymbol{D}_i(\tilde{\boldsymbol{\beta}}^{(m)}) \right)^{-1} \sum_{i=1}^{N} \boldsymbol{D}_i^T(\tilde{\boldsymbol{\beta}}^{(m)}) \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\alpha}, \tilde{\boldsymbol{\beta}}^{(m)})(\mathbf{Z_i} - \boldsymbol{\pi}_i(\tilde{\boldsymbol{\beta}}^{(m)})). \tag{17}$$

In order to estimate the unknown $\boldsymbol{\alpha}$ in above equation, we first construct a matrix $\mathbf{S_i}$ as the product of centralized $\mathbf{Z_i}$ given by:

$$\mathbf{S_i}(\boldsymbol{\beta}) = (\mathbf{Z_i} - \boldsymbol{\pi}_i(\boldsymbol{\beta}))(\mathbf{Z_i} - \boldsymbol{\pi}_i(\boldsymbol{\beta}))'. \tag{18}$$

Notice that E($\mathbf{S_i}(\boldsymbol{\beta})$) is the variance-covariance matrix of $\mathbf{Z_i}$ under the joint model (i.e. $\boldsymbol{\Sigma}_i(\boldsymbol{\alpha},\boldsymbol{\beta})$). Therefore, one can view $\mathbf{S_i}(\boldsymbol{\beta})$ as the "observed outcome" and $\boldsymbol{\Sigma}_i(\boldsymbol{\alpha},\boldsymbol{\beta})$ as the "expected mean" in estimating $\boldsymbol{\alpha}$ given $\boldsymbol{\beta}$ under the joint model. Let $\mathbf{s_i}(\boldsymbol{\beta}) = vech(\mathbf{S_i}(\boldsymbol{\beta}))$[25] 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 $\boldsymbol{\beta}$, one can estimate the correlation parameter $\boldsymbol{\alpha}$ by minimizing sum of the Euclidean norms of $(\mathbf{s}_i(\boldsymbol{\beta}) - \boldsymbol{\sigma}_i(\boldsymbol{\alpha},\boldsymbol{\beta}))$ as:

$$\sum_{i=1}^{N} (\mathbf{s}_i(\boldsymbol{\beta}) - \boldsymbol{\sigma}_i(\boldsymbol{\alpha},\boldsymbol{\beta}))^T (\mathbf{s}_i(\boldsymbol{\beta}) - \boldsymbol{\sigma}_i(\boldsymbol{\alpha},\boldsymbol{\beta})), \tag{19}$$

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

$$\sum_{i=1}^{N} \frac{\partial \boldsymbol{\sigma}_i(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial \boldsymbol{\alpha}}^T \{\mathbf{s}_i(\boldsymbol{\beta}) - \boldsymbol{\sigma}_i(\boldsymbol{\alpha},\boldsymbol{\beta})\} = 0. \tag{20}$$

Now we can carry out the iterative estimation procedure in the following manner:
Step 1: Obtain an initial estimate of $\boldsymbol{\beta}$ denoted as $\hat{\boldsymbol{\beta}}^0$ from fitting separate model to each outcome;
Step 2: Estimate $\hat{\boldsymbol{\alpha}}^0$ by minimizing Equation (19) with $\boldsymbol{\beta}$ replaced by $\hat{\boldsymbol{\beta}}^0$;
Step 3: Solve for $\hat{\boldsymbol{\beta}}^1$ from generalized estimating Equation (11) with $\boldsymbol{\alpha}$ replaced by $\hat{\boldsymbol{\alpha}}^0$.
Step 4: Iterate between Steps 2 and 3 until convergence criteria are fulfilled for both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

Let us denote the solution from Step 4 as $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$. The $\hat{\boldsymbol{\alpha}}$ is obtained by minimizing Equation (19). Therefore, $\hat{\boldsymbol{\alpha}}$ is a least square estimate, which is consistent and asymptotically normal.[26,27] The $\hat{\boldsymbol{\beta}}$ obtained from above iterative procedure is also 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[28] as follows:

$$\hat{cov}(\hat{\boldsymbol{\beta}}) = \mathbf{A_n}^{-1}\mathbf{B_n}\mathbf{A_n}^{-1} \tag{21}$$

$$\text{where } \mathbf{A_n} = \sum_{i=1}^{N} D_i(\hat{\boldsymbol{\beta}})^T \Sigma_i(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})^{-1} D_i(\hat{\boldsymbol{\beta}})$$

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

Therefore, the confidence intervals and hypothesis testing for $\boldsymbol{\beta}$ can be constructed using the Wald method.

## 4. SIMULATION STUDY

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

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

To construct the latent multivariate normal process, we generated a $8 \times 1$ random vector $\boldsymbol{\epsilon_i}$=$(\epsilon_{i10}, \epsilon_{i11}, \epsilon_{i12}, \epsilon_{i13}, \epsilon_{i14}, \epsilon_{i20}, \epsilon_{i22}, \epsilon_{i24})$. $\boldsymbol{\epsilon_i}$ follows a multivariate normal distribution with correlation $\mathbf{R}$. $\mathbf{R}$ is a function of the correlation parameter $\boldsymbol{\alpha} = (\alpha_{11}, \alpha_{22}, \alpha_{12}, \alpha_{212})$. In this simulation study, we constructed $\mathbf{R}$ as either an extended AR-type correlation as in (5) or an extended exchangeable correlation as in (6). After obtaining the marginal probabilities and the latent process $\boldsymbol{\epsilon_i}$, each element of the response vector $\boldsymbol{Y_i}$ can be generated as:

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

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

The above process ensures the response vector $\boldsymbol{Y_i}$ has the specified marginal probabilities given by (1) and the specified correlation based on the latent normal process given by either (5) or (6).

We examined different correlation structures and parameters. For each scenario, $M = 1000$ Monte-Carlo datasets with $n = 100$ and $n = 50$ subjects were generated. We fitted three models to each dataset:

(1) sep-GEE: separately fitting the ordinal outcome and the binary outcome using GEE;

(2) Joint GEE Independence: fitting the proposd joint model with independent correlation structure;

(3) Joint GEE: fitting the joint model where the correlation parameter $\boldsymbol{\alpha}$ and the regression parameter $\boldsymbol{\beta}$ are estimated.

We conducted all analyses using R version 3.3.0. Table 1 shows the simulation results with $n = 100$ subjects. We examined four sets of $\boldsymbol{\alpha}$ representing strong and moderate correlation for the extended AR-type and extended

exchangeable correlation. The regression coefficients are estimated along with their Monte-Carlo variances and robust sandwich variance estimates. Table 1 shows the estimators for regression coefficients are unbiased in all models, because GEE estimators remain unbiased even with misspecified correlation.[28] In addition, the coverage probabilities for all models are close to the nominal level and the robust variance estimates are close to the Monte-Carlo variances. This implies the robust variance estimators are approximately unbiased even with misspecified correlation.

However, when the joint model is adopted with the correct correlation structure, the estimators gain efficiencies, in some cases, by up to $18\%$. Generally, the efficiency gains increase as the correlation increase. More specifically, the efficiency gains of baseline covariates for the binary outcome ($\hat{\beta}_{2x}$) are $18\%$ and $11\%$ in scenarios 1 and 3, and $6\%$ and $1\%$ in scenarios 2 and 4. The estimates of time covariates for both outcomes ($\hat{\beta}_{1t}, \hat{\beta}_{2t}$) have efficiency gains up to 12% and 17% in scenarios 1 and 3, and 3% and 1% in scenarios 2 and 4. The estimate of baseline covariate for the ordinal outcome ($\hat{\beta}_{1x}$) has a slight efficiency gains up to 4%. Supplementary Table 1 lists results with $n = 50$ subjects, which is similar to Table 1, except the standard errors for the estimates are larger due to the smaller sample size.

When the correct correlation structure is specified, Table 2 shows that the joint model yields unbiased estimates for correlation parameters, which may be of further interest for assessing the association among multiple outcomes.

To check the robustness of the proposed joint model for misspecified correlation structure, we generated data with the extended exchangeable and AR-type correlations. In both cases, we fitted the sep-GEE and joint model with extended AR-type and exchangeable correlation. Table 3 shows the performance of the joint model when the working correlation structure is misspecified. The estimators for regression coefficients remain unbiased with coverage probabilities close to the nominal level. However, variance of the estimates is the smallest with the correct correlation structure. And the efficiency gains decrease with misspecified working correlation structure. In some cases, specifically with misspecified correlation structure, the joint model might results in efficiency loss compared to separate modeling. The exchangeable correlation structure appears to be more robust than the independence structure (sep-GEE) and the AR correlation structure in terms of maintaining the coverage and efficiency against misspecification of the true correlation structure.

To investigate higher dimensional situations, we performed an additional simulation study with three longitudinal outcomes (one four-category ordinal outcome with 4 repeated measurements and two binary outcomes with 3 repeated measurements each). The results listed in supplementary Table 2 shows the proposed joint model still has efficiency gains but at the cost of increased computational time.

## 5. VIRAHEP-C STUDY DATA ANALYSIS

The Virahep-C Study is a nonrandomized, multicenter clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). A total of 401 patients, 196 African American (AA) and 205 Caucasian American (CA), were enrolled. All participants were to receive treatments with peginterferon alfa-2a (180$\mu$g/wk) once weekly and ribavirin (1000-1200mg/day) twice daily. In this study, monitors placed on prescription bottles, referred to as MEMS caps (Medication Event Management System, Aardex, Zug, Switzerland), were used to monitor drug consumption. These MEMS caps continuously recorded time when the prescription bottles were closed, which was assumed to be the time drugs were taken. Based on the daily cap closing, patients' adherence to daily ribavirin was categorized as fully adherent (2 or more closings), partially adherent (1 closing), or non-adherent (no closings). Patients' adherence to weekly peginterferon was characterized as fully adherent (at least 1 closing) and non-adherent (no closings). Therefore, each patient's adherence outcome consists of one binary and one ordinal longitudinal responses. Although participants in the Virahep-C study continued treatments for up to 48 weeks, it is believed that adherence during the first 12 weeks is more important to achieve positive responses to therapy.[8] Hence, our analysis primarily focus on the first 12 weeks.

Figure 1 shows the overall trend of adherence to peginterferon and ribavirin fitted using lowess splines for AA and CA patients. On average, patients remain adherent to peginterferon during the first 12 weeks, while adherence to ribavirin

decreases over time. We also notice that CA patients have better adherence for both medications compared to AA patients. There was a sudden drop in adherence to peginterferon during week 4 of the therapy. This drop is an artificial phenomenon because patients were allowed to switch the day of peginterferon shot starting at week 4. Some patients started taking shots during weekends instead of clinic visits, which are generally on weekdays. This resulted in some patients taking no shot during week 4, but more than one shot during week 5.

We first present analysis results from fitting separate modeling using generalized estimating equations.[28] With main covariate of interest, race (CA:1, AA:0), remaining in the model, backward model selection was conducted. Four other covariates with p value $< 0.1$ in the univariate model also remain in the final model including time (number of days since the treatment was initiated), sex (male:1, female:0), baseline log HCV RNA level (vload), employment status (unemployed:1, employed:0) along with time and race interaction. So, the marginal probability model for the ordinal adherence to ribavirin can be expressed as follow:

$$log\left(\frac{Pr(Y_{i1t} \leq k)}{1 - Pr(Y_{i1t} \leq k)}\right) = a_{1k} + \beta_1 \times race_i + \beta_2 \times t + \beta_3 \times race_i * t + \beta_4 \times sex_i + \beta_5 \times vload_i + \beta_6 \times employ_i, \tag{22}$$

where $t = 0, 1, 2, \ldots, 84$ and $k = 0, 1$. Similarly, the marginal probability model for the binary adherence to peginterferon can be expressed as follow:

$$log\left(\frac{Pr(Y_{i2t} \leq 0)}{1 - Pr(Y_{i2t} \leq 0)}\right) = a_{20} + \beta_1' \times race_i + \beta_2' \times t + \beta_3' \times race_i * t + \beta_4' \times sex_i + \beta_5' \times vload_i + \beta_6' \times employ_i, \tag{23}$$

where $t = 0, 7, 14, \ldots, 84$. In above cumulative logistic and logistic models, a positive regression coefficient indicates a higher probability of being in a lower adherence category.

We fitted both marginal models assuming an auto-regressive (AR) correlation structure. Table 4 shows the regression parameter estimates. The main covariate of interest, race, is significantly associated with adherence to both peginterferon and ribavirin. CA patients are more likely to be adherent to both drugs compared to AA patients. This is consistent with previous study findings.[6] Employment status is also significant in peginterferon adherence, but not ribavirin. Finally, the coefficients of time for both outcomes are positive indicating patients became less adherent to their medications over time.

Even though it seems that separate modeling serves the purpose of finding predictors for individual medication adherence outcome, it fails to account for the correlation between two longitudinal outcomes. The joint modeling approach not only addresses this, but also provides an estimate of the correlation between two outcomes.

Therefore, we analyzed the data using the joint modeling approach. We assumed the same marginal models for both longitudinal outcomes and further assumed an extended AR-type correlation between any pair of observations from the same subject. More specifically, for each subject $i$, we defined the correlation structure as follows:

$$\rho_{ijt,ij't'} = \begin{cases} \alpha_{11}^{|t-t'|} & \text{if } j = j' = 1 \\ \alpha_{22}^{|t-t'|} & \text{if } j = j' = 2 \\ \alpha_{12}^{|t-t'|} \times \alpha_{212} & \text{otherwise ,} \end{cases}$$

where $\alpha_{11}$ represents the correlation between two ribavirin adherence events observed one day apart; $\alpha_{22}$ represents corresponding correlation between two peginterferon adherence events; and $\alpha_{12}^{\Delta t} \times \alpha_{212}$ represents the correlation between ribavirin adherence and peginterferon adherence observed $\Delta t$ days apart.

The correlation estimates from the joint model are $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_{11}, \hat{\alpha}_{22}, \hat{\alpha}_{12}, \hat{\alpha}_{212})$=$(0.20, 0.98, 0.99, 0.42)$. This suggests that adherence to weekly peginterferon has much higher within-outcome correlation compared to daily ribavirin. Since $\hat{\alpha}_{12} = 0.99$ is close to 1, it suggests that between outcome correlation does not decrease significantly as the length of time

increases.

The regression estimates from the joint model are shown in Table 5. The effect of race is no longer statistically significant for ribavirin adherence, and the p-value of race for peginterferon adherence is 0.053. Adherence has a statistically significant decline over time for ribavirin but not for peginterferon. And employment status is the only patient characteristic has a significant effect on the adherence to peginterferon. The odds ratio of being nonadherent to peginterferon is $exp(0.81) = 2.25$ for a patient who is unemployed at baseline compared to an otherwise similar patient who is employed at baseline.

## 6. DISCUSSION

Multivariate ordinal longitudinal outcome (e.g. medication adherence, patient reported quality of life outcome, etc.) is common in biomedical research. In this paper, we propose a joint model with flexible correlation structure, which addresses the two main challenges to analyze this type of data using GEE based methods. First, we use a latent multivariate normal process to link each longitudinal outcome together and used the polychoric correlation as an association measure. Second, we construct a flexible framework to model the multilevel correlation structure. The proposed joint model accounts for both the between and within outcome association via the correlation of the latent multivariate normal process. Such correlation is not restricted by the marginal probabilities. The simulation study indicates that estimators for the joint model are unbiased and more efficient compared to those obtained from fitting separate GEEs when the correlation structure is correctly specified. However, they may be less efficient in some cases with misspecified correlation structure. As observed in the analysis results of the Virahep-C study, failure to account for the between outcome correlation can lead to different conclusions.

In the proposed joint model, although we assumed the observed ordinal outcomes arose from a latent multivariate normal process, as argued in Qu et al.,[21] this assumption does not affect the consistency of $\hat{\boldsymbol{\beta}}$ and the estimation of marginal probabilities. This is because the assumed latent normal process affects only the correlation parameters through the pairwise joint probabilities, but not the marginal probabilities. Thus, even if the latent normal process is misspecified, $\hat{\boldsymbol{\beta}}$ will remain unbiased, while $\hat{\boldsymbol{\alpha}}$ can be viewed as approximations for the true correlation.

We demonstrated in Section 3.3 that the joint model can incorporate a flexible multilevel correlation structure as extended AR-type (5), extended exchangeable (6), and extended m-dependent (7). One can construct even more flexible correlation structures by combining different correlation types. For example, one can assume the measurements from the same outcome has an AR(1) correlation $\rho_{jt,jt'} = \alpha_{jj}^{|t-t'|}$, while the measurements from different outcomes have either an extended exchangeable correlation $\rho_{ijt,ij't'} = \alpha_{2jj'}$ or an extended m-dependent correlation $\rho_{ijt,ij't'} = \alpha_{2jj'|t-t'|}$. These two mixed correlation structures can be written as:

$$\rho_{ijt,ij't'} = \alpha_{jj'}^{|t-t'|\times I(j=j')} \times \alpha_{2jj'}^{I(j\neq j')}, \text{ and} \tag{24}$$

$$\rho_{ijt,ij't'} = \alpha_{jj'}^{|t-t'|\times I(j=j')} \times \alpha_{2jj'|t-t'|}^{I(j\neq j')}. \tag{25}$$

One can also extend it to handle more than two levels of association (e.g. association due to sites). Although additional correlation parameters are required, the estimation process remain the same. Because the proposed joint model uses the least square method to estimate correlation parameter $\boldsymbol{\alpha}$, it provides consistent estimators for $\boldsymbol{\alpha}$. Unlike Liang and Zeger[28] only provided formulas to estimate $\boldsymbol{\alpha}$ for several simple correlation structures (e.g. exchangeable, AR, m-dependent).

We also investigated the performance of the proposed joint model in higher dimensional situations. In the simulation scenarios considered in Section 4, joint modeling still resulted in efficiency gains that comes at the cost of increased computing time. For example, we analyzed a dataset with sample size of 150, where each subject has one four-category ordinal outcome with 4 repeated measurements and two binary outcomes with 3 repeated measurements each

(Supplementary Table 2). Using the R codes provided with this paper, three separate GEE runs used approximately 2 seconds in total compared to 1200 seconds for the joint model.

In the Virahep-C Study, adherence data were collected via digital caps, and hence data were rarely missing. However, missing data are usually inevitable for most longitudinal outcomes. Standard GEE inference is valid only when data are complete or missing completely at random (MCAR). The proposed estimators will be biased if data are missing at random (MAR). There are two general approaches to obtain consistent estimators with MAR data under GEE framework: multiple imputation[29,30] and inverse probability weighting.[31] The idea of multiple imputation is to fill in imputed data and then apply standard GEE methods. Multiple imputation can be implemented by commonly available softwares, such as PROC MI and PROC MIANALYZE in SAS. The inverse probability weighting (IPW) method was first introduced in survey studies.[32] Robins et al.[31] proposed a class of IPW estimating equation, which extends the standard GEE method to MAR data. Later, Yi and Cook[33] proposed a weighted second-order estimating equations, where the inverse probability weighting was applied to both mean and association parameters. Therefore, it provides consistent estimators for both marginal mean parameters and association parameters. Such weighted second-order estimating equation can be implemented to the proposed joint model in dealing with MAR data. This is beyond the scope of this paper and will be presented in a separate paper.

## References

1. Osterberg L, Blaschke T. Adherence to Medication. *N Engl J Med.* 2005;353(5):487-497.
2. Horwitz RI, Horwitz SM. Adherence to treatment and health outcomes. *Arch Intern Med.* 1993;153(16):1863-1868.
3. LaRosa JC. Poor compliance: the hidden risk factor. *Curr Atheroscler Rep.* 2000;2(1):1-4.
4. Waeber B, Leonetti G, Kolloch R, McInnes GT. Compliance with aspirin or placebo in the Hypertension Optimal Treatment(HOT) study. *J Hypertens.* 1999;17(7):1041-1045.
5. Nose M, Barbui C, Gray R, Tansella M. Clinical interventions for treatment nonadherence in psychosis: meta-analysis. *Br J Psychiatry.* 2003;183:197-206.
6. Conjeevaram HS, Fried MW, Jeffers LJ, et al. Peginterferon and ribavirin treatment in African American patients and Caucasian American patients with hepatitis C genotype 1. *Gastroenterology.* 2006;131(2):470-477.
7. Raptopoulou M, Tsantoulas D, Vafiadi I, et al. 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. *J Viral Hepat.* 2005;12(1):91-95.
8. Shiffman F, Bacon BR, Nelson D, et al. Peginterferon alfa-2a and ribavirin for 16 or 24 weeks in HCV genotype 2 or 3. *N Engl J Med.* 2007;357(2):124-134.
9. Verveke G, Davidian M. *Longitudinal Data Analysis*. CRC Press; 2009.
10. Molenberghs G, Verbeke G. *Models for Discrete Longitudinal Data*. Springer; 2005.
11. Molenberghs G, Geys H, Buyse M. Evaluation of surrogate endpoints in randomized experiments with mixed discrete and continuous outcomes. *Stat Med.* 2001;20(20):3023-3038.
12. Prentice RL, Zhao LP. Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics.* 1991;47(3):825-839.
13. Tate RF. Correlation between a discrete and a continuous variable. *Ann Math Statist.* 1954;25(3):603-607.
14. Little RJA, Schluchter MD. Maximum likelihood estimation for mixed continuous and categorical data with missing values. *Biometrika.* 1985;72(3):497-512.
15. Prentice RL. Correlated binary regression with covariates specific to each binary observation. *Biometrics.* 1988;44(4):1033-1048.
16. Liang KY, Zeger SL, Qaqish B. Multivariate regression analyses for categorical data. *J R Stat Soc Series B Stat Methodol.* 1992;54(1):3-40.
17. Carey V, Zeger SL, Diggle P. Modelling multivariate binary data with alternating logistic regressions. *Biometrika.* 1993;80(3):517-526.
18. Qu Y, Williams GW, Beck GJ, Medendorp SV. Latent variable models for clustered dichotomous data with multiple subclusters. *Biometrics.* 1992;48(4):1095-1102.
19. Miller ME, Davis CS, Landis JR. The analysis of longitudinal polytomous data: Generalized estimating equations and connections with weighted least squares. *Biometrics.* 1993;49(4):1033-1044.
20. Williamson JM, Kim K, Lipsitz SR. Analyzing bivariate ordinal data using a global odds ratio. *J Am Stat Assoc.* 1995;90(432):1432-1437.
21. Qu Y, Piedmonte MR, Medendorp SV. Latent Variable Models for Clustered Ordinal Data. *Biometrics.* 1995;51(1):268-275.

22. Lipsitz SR, Fitzmaurice GM, Ibrahim JG, Sinha D, Parzen M, Lipshultz S. Joint generalized estimating equations for multivariate longitudinal binary outcomes with missing data: an application to acquired immune deficiency syndrome data. *J R Stat Soc Ser A Stat Soc.* 2009;172(1):3-20.
23. Sutradhar BC, Kovacevic M. Analysing ordinal longitudinal survey data: Generalised estimating equations approach. *Biometrika.* 2000;87(4):837-848.
24. Harville DA, Mee RW. A mixed model procedure for analyzing ordered categorical data. *Biometrics.* 1984;40(2):393-408.
25. Vonesh EF, Chinchilli VM. *Linear and nonlinear models for the analysis of repeated measurements*. CRC Press; 1996.
26. Jennrich R. Asymptotic properties of non-linear least square estimators. *Ann Math Statist.* 1969;40(2):633-643.
27. Eicker F. Asymptotic normality and consistency of the least squares estimators for family of linear regression. *Ann Math Statist.* 1963;34(2):447-456.
28. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika.* 1986;73(1):13-22.
29. Rubin DB. Multiple Imputations in Sample Surveys: A Phenomenological Bayesian Approach to Nonresponse. *Proc Am Stat Assoc.* 1978;:20-34.
30. Beunckens C, Sotto C, Molenberghs G. A simulation study comparing weighted estimating equations with multiple imputation based estimating equations for longitudinal binary data. *Comput Stat Data Anal.* 2008;52(3):1533-1548.
31. Robins JM, Rotnitzky A, Zhao LP. Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data. *J Am Stat Assoc.* 1995;90(429):106-121.
32. Horvitz DG, Thompson DJ. A Generalization of sampling without replacement from a finite universe. *J Am Stat Assoc.* 1952;47:663-685.
33. Yi GY, Cook RJ. Marginal Methods for Incomplete Longitudinal Data Arising in Clusters. *J Am Stat Assoc.* 2002;97(450):1071-1080.

**Table 1.** Performance of different methods with extended exchangeable and AR-type correlation. Results are from M=1000 datasets with n=100 subjects. $\hat{\boldsymbol{\beta}}$ is the mean of Monte-Carlo (MC) estimates; var($\hat{\boldsymbol{\beta}}$) is the MC variance of $\hat{\boldsymbol{\beta}}$; $\hat{var}(\hat{\boldsymbol{\beta}})$ is the MC mean of estimated variances; CP% is the empirical coverage probability of 95% Wald confidence interval; RE is the relative efficiency of estimator based on Joint GEE[c] compared to sep-GEE[a].

| | | | sep-GEE[a] | | | | Joint GEE Independence[b] | | | | Joint GEE[c] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| True Correlation | Para-meters | $\boldsymbol{\beta}$ | $\hat{\boldsymbol{\beta}}$ | var($\hat{\boldsymbol{\beta}}$) | $\hat{var}(\hat{\boldsymbol{\beta}})$ | CP% | $\hat{\boldsymbol{\beta}}$ | var($\hat{\boldsymbol{\beta}}$) | $\hat{var}(\hat{\boldsymbol{\beta}})$ | CP% | $\hat{\boldsymbol{\beta}}$ | var($\hat{\boldsymbol{\beta}}$) | $\hat{var}(\hat{\boldsymbol{\beta}})$ | CP% | RE |
| 1) | $\beta_{1x}$ | 0.3 | 0.30 | 0.145 | 0.133 | 95% | 0.30 | 0.145 | 0.133 | 95% | 0.30 | 0.143 | 0.131 | 94% | 1.01 |
| Exchangeable | $\beta_{1t}$ | 0.2 | 0.20 | 0.002 | 0.001 | 95% | 0.20 | 0.001 | 0.001 | 95% | 0.20 | 0.001 | 0.001 | 95% | 1.07 |
| $(\alpha_{11}, \alpha_{22}, \alpha_{12})$ | $\beta_{2x}$ | 0.4 | 0.41 | 0.211 | 0.192 | 95% | 0.41 | 0.231 | 0.206 | 94% | 0.41 | 0.179 | 0.171 | 95% | 1.18 |
| $= (0.9, 0.9, 0.9)$ | $\beta_{2t}$ | 0.5 | 0.52 | 0.011 | 0.010 | 94% | 0.52 | 0.011 | 0.010 | 94% | 0.52 | 0.010 | 0.009 | 94% | 1.17 |
| 2) | $\beta_{1x}$ | 0.3 | 0.30 | 0.128 | 0.117 | 94% | 0.30 | 0.128 | 0.117 | 94% | 0.31 | 0.125 | 0.114 | 94% | 1.02 |
| Exchangeable | $\beta_{1t}$ | 0.2 | 0.20 | 0.002 | 0.002 | 95% | 0.20 | 0.002 | 0.002 | 95% | 0.20 | 0.002 | 0.002 | 95% | 1.00 |
| $(\alpha_{11}, \alpha_{22}, \alpha_{12})$ | $\beta_{2x}$ | 0.4 | 0.41 | 0.184 | 0.174 | 94% | 0.40 | 0.193 | 0.179 | 93% | 0.40 | 0.174 | 0.170 | 94% | 1.06 |
| $= (0.8, 0.7, 0.4)$ | $\beta_{2t}$ | 0.5 | 0.51 | 0.010 | 0.010 | 94% | 0.51 | 0.010 | 0.010 | 94% | 0.51 | 0.010 | 0.010 | 94% | 1.00 |
| 3) | $\beta_{1x}$ | 0.3 | 0.30 | 0.122 | 0.114 | 95% | 0.30 | 0.129 | 0.121 | 95% | 0.30 | 0.121 | 0.116 | 95% | 1.01 |
| AR-type | $\beta_{1t}$ | 0.2 | 0.20 | 0.003 | 0.003 | 95% | 0.20 | 0.003 | 0.003 | 96% | 0.20 | 0.003 | 0.003 | 96% | 1.12 |
| $(\alpha_{11}, \alpha_{22}, \alpha_{12}, \alpha_{212})$ | $\beta_{2x}$ | 0.4 | 0.42 | 0.198 | 0.183 | 96% | 0.41 | 0.207 | 0.191 | 95% | 0.41 | 0.178 | 0.201 | 95% | 1.11 |
| $= (0.9, 0.9, 0.9, 0.9)$ | $\beta_{2t}$ | 0.5 | 0.52 | 0.011 | 0.010 | 94% | 0.52 | 0.011 | 0.010 | 94% | 0.51 | 0.010 | 0.010 | 94% | 1.06 |
| 4) | $\beta_{1x}$ | 0.3 | 0.31 | 0.109 | 0.100 | 94% | 0.31 | 0.109 | 0.100 | 94% | 0.30 | 0.104 | 0.096 | 95% | 1.04 |
| AR-type | $\beta_{1t}$ | 0.2 | 0.20 | 0.004 | 0.004 | 96% | 0.20 | 0.004 | 0.004 | 96% | 0.20 | 0.003 | 0.004 | 96% | 1.03 |
| $(\alpha_{11}, \alpha_{22}, \alpha_{12}, \alpha_{212})$ | $\beta_{2x}$ | 0.4 | 0.41 | 0.153 | 0.147 | 95% | 0.41 | 0.156 | 0.148 | 94% | 0.41 | 0.151 | 0.145 | 95% | 1.01 |
| $= (0.8, 0.7, 0.4, 0.3)$ | $\beta_{2t}$ | 0.5 | 0.51 | 0.013 | 0.012 | 94% | 0.51 | 0.013 | 0.012 | 94% | 0.51 | 0.012 | 0.012 | 94% | 1.01 |

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

**Table 2.** Estimates and Monte-Carlo variances of correlation parameters in Tables 1 and Supplementary Table 1 estimated from the joint GEE. A) Table 1: M=1000 datasets with n=100 subjects; B) Supplementary Table 1: M=1000 datasets with n=50 subjects. $\hat{\boldsymbol{\alpha}}$ is the mean of Monte-Carlo (MC) estimates; $var(\hat{\boldsymbol{\alpha}})$ is the MC variance of $\hat{\boldsymbol{\alpha}}$.

| Correlation | Correlation | | A | | B | |
|---|---|---|---|---|---|---|
| Type | Parameters | $\boldsymbol{\alpha}$ | $\hat{\boldsymbol{\alpha}}$ | $var(\hat{\boldsymbol{\alpha}})$ | $\hat{\boldsymbol{\alpha}}$ | $var(\hat{\boldsymbol{\alpha}})$ |
| Exchangeable | $\alpha_{11}$ | 0.9 | 0.94 | 0.004 | 0.89 | 0.004 |
| | $\alpha_{22}$ | 0.9 | 0.86 | 0.003 | 0.90 | 0.012 |
| | $\alpha_{12}$ | 0.9 | 0.90 | 0.002 | 0.90 | 0.003 |
| | $\alpha_{11}$ | 0.8 | 0.79 | 0.003 | 0.78 | 0.006 |
| | $\alpha_{22}$ | 0.7 | 0.60 | 0.008 | 0.68 | 0.030 |
| | $\alpha_{12}$ | 0.4 | 0.40 | 0.014 | 0.41 | 0.025 |
| | $\alpha_{11}$ | 0 | -0.00 | 0.005 | -0.01 | 0.005 |
| | $\alpha_{22}$ | 0 | 0.04 | 0.013 | -0.03 | 0.051 |
| | $\alpha_{12}$ | 0 | 0.00 | 0.006 | 0.01 | 0.007 |
| AR-type | $\alpha_{11}$ | 0.9 | 0.91 | 0.002 | 0.89 | 0.002 |
| | $\alpha_{22}$ | 0.9 | 0.84 | 0.025 | 0.89 | 0.024 |
| | $\alpha_{12}$ | 0.9 | 0.89 | 0.004 | 0.89 | 0.003 |
| | $\alpha_{212}$ | 0.9 | 0.90 | 0.003 | 0.90 | 0.005 |
| | $\alpha_{11}$ | 0.8 | 0.79 | 0.002 | 0.78 | 0.005 |
| | $\alpha_{22}$ | 0.7 | 0.60 | 0.009 | 0.65 | 0.047 |
| | $\alpha_{12}$ | 0.4 | 0.13 | 0.005 | 0.37 | 0.174 |
| | $\alpha_{212}$ | 0.3 | 0.24 | 0.015 | 0.31 | 0.037 |
| | $\alpha_{11}$ | 0 | -0.01 | 0.007 | -0.01 | 0.014 |
| | $\alpha_{22}$ | 0 | 0.00 | 0.001 | 0.06 | 0.026 |
| | $\alpha_{12}$ | 0 | -0.01 | 0.032 | 0.02 | 0.112 |
| | $\alpha_{212}$ | 0 | -0.00 | 0.017 | 0.00 | 0.035 |

**Table 3.** Performance of the joint model when correlation structure is misspecified. Results are from M=500 datasets with n=100 subjects. $\hat{\boldsymbol{\beta}}$ is the mean of Monte-Carlo (MC) estimates; var($\hat{\boldsymbol{\beta}}$) is the MC variance of $\hat{\boldsymbol{\beta}}$; $\hat{var}(\hat{\boldsymbol{\beta}})$ is the MC mean of estimated variances; CP% is the empirical coverage probability of 95% Wald confidence interval.

| | | | sep-GEE[a] | | | | Joint GEE Exchangeable[b] | | | | | Joint GEE AR[c] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| True Correlation | Para-meters | $\boldsymbol{\beta}$ | $\hat{\beta}$ | var($\hat{\beta}$) | $\hat{var}(\hat{\boldsymbol{\beta}})$ | CP% | $\hat{\beta}$ | var($\hat{\beta}$) | $\hat{var}(\hat{\boldsymbol{\beta}})$ | CP% | RE | $\hat{\beta}$ | var($\hat{\beta}$) | $\hat{var}(\hat{\boldsymbol{\beta}})$ | CP% | RE |
| | $\beta_{1x}$ | 0.3 | 0.32 | 0.113 | 0.101 | 94% | 0.32 | 0.112 | 0.099 | 94% | 1.02 | 0.32 | 0.109 | 0.096 | 95% | 1.04 |
| AR-type | $\beta_{1t}$ | 0.2 | 0.20 | 0.004 | 0.004 | 95% | 0.20 | 0.004 | 0.004 | 95% | 1.00 | 0.20 | 0.004 | 0.004 | 95% | 1.00 |
| $(\alpha_{11}, \alpha_{22}, \alpha_{12}, \alpha_{212})$ | $\beta_{2x}$ | 0.4 | 0.41 | 0.163 | 0.150 | 93% | 0.41 | 0.156 | 0.147 | 95% | 1.05 | 0.41 | 0.155 | 0.144 | 94% | 1.05 |
| $= (0.8, 0.7, 0.4, 0.3)$ | $\beta_{2t}$ | 0.5 | 0.51 | 0.013 | 0.012 | 93% | 0.51 | 0.013 | 0.012 | 94% | 1.01 | 0.51 | 0.013 | 0.012 | 94% | 1.01 |
| | $\beta_{1x}$ | 0.3 | 0.31 | 0.131 | 0.118 | 93% | 0.31 | 0.129 | 0.114 | 95% | 1.02 | 0.31 | 0.132 | 0.117 | 95% | 0.99 |
| Exchangeable | $\beta_{1t}$ | 0.2 | 0.20 | 0.002 | 0.002 | 94% | 0.20 | 0.002 | 0.002 | 94% | 1.06 | 0.20 | 0.002 | 0.002 | 94% | 0.95 |
| $(\alpha_{11}, \alpha_{22}, \alpha_{12})$ | $\beta_{2x}$ | 0.4 | 0.41 | 0.175 | 0.173 | 96% | 0.41 | 0.164 | 0.169 | 96% | 1.07 | 0.41 | 0.164 | 0.170 | 96% | 1.07 |
| $= (0.8, 0.7, 0.4)$ | $\beta_{2t}$ | 0.5 | 0.51 | 0.010 | 0.010 | 93% | 0.51 | 0.010 | 0.010 | 94% | 1.01 | 0.51 | 0.010 | 0.010 | 94% | 1.01 |

sep-GEE[a]: seperately fit the ordinal outcome and binary outcome using GEE; 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 4.** Regression estimates (standard errors) and p values from separate regression models, for Peginterferon and Ribavirin adherence.

| Covariates | Peginterferon | | Ribavirin | |
|---|---|---|---|---|
| | Est(SE) | p value | Est(SE) | p value |
| Race(CA vs. AA) | -0.66(0.298) | 0.027 | -0.56(0.205) | 0.006 |
| Time(days) | 0.01(0.003) | 0.033 | 0.01(0.002) | $< .001$ |
| Race*Time(days) | -0.00(0.005) | 0.892 | -0.00(0.003) | 0.509 |
| Sex (male vs. female) | -0.11(0.284) | 0.692 | 0.02(0.190) | 0.902 |
| Baseline viral load($log_{10}$) | 0.09(0.145) | 0.547 | 0.07(0.115) | 0.556 |
| Employed (no vs. yes) | 0.88(0.273) | 0.001 | -0.06(0.186) | 0.767 |

**Table 5.** Regression estimates (standard errors) and p values from the joint model for Peginterferon and Ribavirin.

| Covariates | Peginterferon | | Ribavirin | |
|---|---|---|---|---|
| | Est(SE) | p value | Est(SE) | p value |
| Race(CA vs. AA) | -0.57(0.294) | 0.053 | -0.32(0.210) | 0.124 |
| Time(days) | 0.00(0.003) | 0.162 | 0.01(0.002) | $< .001$ |
| Race*Time(days) | -0.00(0.005) | 0.893 | -0.00(0.003) | 0.194 |
| Sex(male vs. female) | -0.06(0.282) | 0.819 | 0.12(0.189) | 0.516 |
| Baseline viral load($log_{10}$) | 0.05(0.151) | 0.742 | 0.02(0.109) | 0.837 |
| Employed(no vs. yes) | 0.81(0.267) | 0.002 | -0.22(0.158) | 0.160 |

Supplementary Table 1: Performance of different methods with extended exchangeable and AR-type correlation. Results are from M=1000 datasets with n=50 subjects. $\hat{\boldsymbol{\beta}}$ is the mean of Monte-Carlo (MC) estimates; var($\hat{\boldsymbol{\beta}}$) is the MC variance of $\hat{\boldsymbol{\beta}}$; $\hat{var}(\hat{\boldsymbol{\beta}})$ is the MC mean of estimated variances; CP% is the empirical coverage probability of 95% Wald confidence interval; RE is the relative efficiency of estimator based on Joint GEE[c] compared to sep-GEE[a].

| | | | sep-GEE[a] | | | | Joint GEE Independence[b] | | | | Joint GEE[c] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| True Correlation | Para-meters | $\boldsymbol{\beta}$ | $\hat{\boldsymbol{\beta}}$ | var($\hat{\boldsymbol{\beta}}$) | $\hat{var}(\hat{\boldsymbol{\beta}})$ | CP% | $\hat{\boldsymbol{\beta}}$ | var($\hat{\boldsymbol{\beta}}$) | $\hat{var}(\hat{\boldsymbol{\beta}})$ | CP% | $\hat{\boldsymbol{\beta}}$ | var($\hat{\boldsymbol{\beta}}$) | $\hat{var}(\hat{\boldsymbol{\beta}})$ | CP% | RE |
| 1) | $\beta_{1x}$ | 0.3 | 0.30 | 0.301 | 0.273 | 94% | 0.30 | 0.30 | 0.272 | 94% | 0.30 | 0.299 | 0.265 | 94% | 1.00 |
| Exchangeable | $\beta_{1t}$ | 0.2 | 0.21 | 0.003 | 0.003 | 94% | 0.21 | 0.003 | 0.003 | 94% | 0.21 | 0.003 | 0.003 | 95% | 1.04 |
| $(\alpha_{11}, \alpha_{22}, \alpha_{12})$ | $\beta_{2x}$ | 0.4 | 0.43 | 0.488 | 0.429 | 95% | 0.43 | 0.492 | 0.427 | 94% | 0.42 | 0.393 | 0.349 | 95% | 1.24 |
| $= (0.9, 0.9, 0.9)$ | $\beta_{2t}$ | 0.5 | 0.53 | 0.022 | 0.021 | 93% | 0.53 | 0.022 | 0.020 | 93% | 0.53 | 0.020 | 0.019 | 94% | 1.09 |
| 2) | $\beta_{1x}$ | 0.3 | 0.28 | 0.259 | 0.237 | 94% | 0.29 | 0.259 | 0.237 | 94% | 0.29 | 0.254 | 0.232 | 95% | 1.02 |
| Exchangeable | $\beta_{1t}$ | 0.2 | 0.21 | 0.004 | 0.004 | 94% | 0.21 | 0.004 | 0.004 | 94% | 0.21 | 0.004 | 0.004 | 93% | 1.02 |
| $(\alpha_{11}, \alpha_{22}, \alpha_{12})$ | $\beta_{2x}$ | 0.4 | 0.44 | 0.399 | 0.367 | 94% | 0.44 | 0.401 | 0.370 | 93% | 0.43 | 0.365 | 0.353 | 95% | 1.10 |
| $= (0.8, 0.7, 0.4)$ | $\beta_{2t}$ | 0.5 | 0.53 | 0.023 | 0.021 | 93% | 0.53 | 0.023 | 0.020 | 93% | 0.53 | 0.022 | 0.021 | 94% | 1.03 |
| 3) | $\beta_{1x}$ | 0.3 | 0.29 | 0.268 | 0.246 | 95% | 0.29 | 0.268 | 0.246 | 95% | 0.29 | 0.269 | 0.237 | 93% | 1.00 |
| AR-type | $\beta_{1t}$ | 0.2 | 0.21 | 0.006 | 0.006 | 94% | 0.21 | 0.006 | 0.006 | 94% | 0.21 | 0.006 | 0.006 | 94% | 1.03 |
| $(\alpha_{11}, \alpha_{22}, \alpha_{12}, \alpha_{212})$ | $\beta_{2x}$ | 0.4 | 0.43 | 0.436 | 0.385 | 94% | 0.43 | 0.453 | 0.395 | 94% | 0.39 | 0.374 | 0.334 | 94% | 1.17 |
| $= (0.9, 0.9, 0.9, 0.9)$ | $\beta_{2t}$ | 0.5 | 0.53 | 0.025 | 0.022 | 93% | 0.53 | 0.026 | 0.021 | 91% | 0.53 | 0.023 | 0.020 | 93% | 1.10 |
| 4) | $\beta_{1x}$ | 0.3 | 0.28 | 0.229 | 0.201 | 94% | 0.28 | 0.231 | 0.201 | 94% | 0.28 | 0.228 | 0.193 | 93% | 1.01 |
| AR-type | $\beta_{1t}$ | 0.2 | 0.21 | 0.008 | 0.008 | 94% | 0.21 | 0.008 | 0.008 | 94% | 0.21 | 0.008 | 0.007 | 94% | 1.03 |
| $(\alpha_{11}, \alpha_{22}, \alpha_{12}, \alpha_{212})$ | $\beta_{2x}$ | 0.4 | 0.44 | 0.310 | 0.310 | 95% | 0.44 | 0.315 | 0.308 | 96% | 0.43 | 0.299 | 0.299 | 96% | 1.04 |
| $= (0.8, 0.7, 0.4, 0.3)$ | $\beta_{2t}$ | 0.5 | 0.52 | 0.028 | 0.026 | 94% | 0.53 | 0.029 | 0.025 | 93% | 0.52 | 0.028 | 0.025 | 94% | 1.01 |

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

Supplementary Table 2: Performance of different methods for higher dimensional outcomes (one four-category ordinal outcome with 4 repeated measurements and two binary outcomes with 3 repeated measurements each) with extended AR-type correlation $(\alpha_{11}, \alpha_{22}, \alpha_{33}, \alpha_{12}, \alpha_{13}, \alpha_{23}, \alpha_{212}, \alpha_{213}, \alpha_{223}) = (0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9)$ . Results are from M=1000 datasets with n=150 subjects. $\hat{\boldsymbol{\beta}}$ is the mean of Monte-Carlo (MC) estimates; var($\hat{\boldsymbol{\beta}}$) is the MC variance of $\hat{\boldsymbol{\beta}}$; $\hat{var}(\hat{\boldsymbol{\beta}})$ is the MC mean of estimated variances; CP% is the empirical coverage probability of 95% Wald confidence interval; RE is the relative efficiency of estimator based on Joint GEE$^c$ compared to sep-GEE$^b$.

| | | sep-GLM$^a$ | | | | sep-GEE$^b$ | | | | Joint GEE$^c$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Parameters | $\boldsymbol{\beta}$ | $\hat{\boldsymbol{\beta}}$ | var($\hat{\boldsymbol{\beta}}$) | $\hat{var}(\hat{\boldsymbol{\beta}})$ | CP% | $\hat{\boldsymbol{\beta}}$ | var($\hat{\boldsymbol{\beta}}$) | $\hat{var}(\hat{\boldsymbol{\beta}})$ | CP% | $\hat{\boldsymbol{\beta}}$ | var($\hat{\boldsymbol{\beta}}$) | $\hat{var}(\hat{\boldsymbol{\beta}})$ | CP% | RE |
| $\beta_{1x}$ | 0.1 | 0.09 | 0.0674 | 0.0218 | 73% | 0.09 | 0.0674 | 0.0715 | 96% | 0.09 | 0.0666 | 0.0708 | 96% | 1.01 |
| $\beta_{1t}$ | 0.2 | 0.20 | 0.0020 | 0.0044 | 100% | 0.20 | 0.0019 | 0.0018 | 94% | 0.21 | 0.0019 | 0.0018 | 95% | 1.02 |
| $\beta_{2x}$ | 0.4 | 0.39 | 0.0910 | 0.0490 | 85% | 0.39 | 0.0883 | 0.0882 | 95% | 0.39 | 0.0845 | 0.0866 | 95% | 1.04 |
| $\beta_{2t}$ | 0.5 | 0.51 | 0.0045 | 0.0075 | 99% | 0.51 | 0.0048 | 0.0048 | 95% | 0.51 | 0.0043 | 0.0044 | 95% | 1.11 |
| $\beta_{3x}$ | 0.3 | 0.29 | 0.0996 | 0.0450 | 81% | 0.30 | 0.0988 | 0.1028 | 96% | 0.30 | 0.0903 | 0.0936 | 94% | 1.09 |
| $\beta_{3t}$ | 0.2 | 0.20 | 0.0043 | 0.0071 | 98% | 0.20 | 0.0044 | 0.0044 | 95% | 0.20 | 0.0039 | 0.0038 | 94% | 1.13 |

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 using GEE with exchangeable correlation and the binary outcome using GEE with AR1 correlation; Joint GEE$^c$: GEE applied to the joint model where correlation parameter $\boldsymbol{\alpha}$ is estimated along with regression parameter $\boldsymbol{\beta}$.