UW Biostatistics Working Paper Series

1-21-2011

Doubly Robust Estimates for Binary Longitudinal Data Analysis with Missing Response and Missing Covariates

Baojiang Chen
University of Washington, bjchen@u.washington.edu

Xiao-Hua Zhou
University of Washington, azhou@u.washington.edu

Doubly Robust Estimates for Binary Longitudinal Data Analysis with Missing Response and Missing Covariates

Baojiang Chen * and Xiao-Hua Zhou

Department of Biostatistics, University of Washington

Seattle, Washington, US 98195

*Email: bjchen@u.washington.edu

Doubly Robust Estimates for Binary Longitudinal Data Analysis with Missing Response and Missing Covariates

Abstract

Longitudinal studies often feature incomplete response and covariate data. Likelihood based method such as EM algorithm gives consistent estimates for data are missing at random provided that the response model and the missing covariate model are correctly specified; while we can misspecify (or do not even to estimate) the distribution of the missing indicators. An alternative method is the weighted estimating equation which gives consistent estimates if the missing data models and response models are correctly specified; but we can misspecify (or do not even to estimate) the distribution of the missing covariates. In this paper we develop a doubly robust estimate method for longitudinal data with missing response and missing covariate when data are missing at random. This method is appealing in that it can provide consistent estimates if either the missing data model or the missing covariate model is correctly specified. Simulation studies demonstrate that this method performs well in a variety of situations.

KEYWORDS: Doubly robust; estimating equation; missing at random; missing covariate; missing response.

1 Introduction

Incomplete longitudinal data often arise in comparative studies because of difficulties in ascertaining responses at scheduled assessment times, partially completed forms or questionnaires, patients refusal to undergo complete examinations, or study subjects failing to attend a scheduled clinic visit. Problems ensue if the mechanism leading to the missing data is dependent on the response or covariates. Analyses based only on individuals with complete data can lead to invalid inferences in this case. Under a missing completely at random (MCAR) mechanism (Little & Rubin, 2002), analyses based on generalized estimating equations (GEE) (Liang & Zeger, 1986) yield consistent estimates of the regression parameters. However, when the data are missing at random (MAR) or missing not at random (MNAR) (Little & Rubin, 2002), analyses based on GEE give inconsistent estimates. Robins et al. (1995) developed a class of inverse probability weighted generalized estimating equations (IPWGEE) which can yield consistent estimates when data are MAR. The weights are obtained from models for the missing data process, and these models must be correctly specified for the resulting estimators to be consistent. Alternatively, one can use maximum likelihood method to estimate the parameters, and it gives consistent estimate if the model is correctly specified.

The literature on methods for missing data has primarily addressed either missing response or missing covariate data (see, e.g., Fitzmaurice et al., 2001; Horton & Laird, 1998; Ibrahim et al., 2001; Lipsitz et al., 1999; Zhao et al., 1996), but relatively little work has been done when both can be missing. In practice, of course, data are often unavailable for both responses and covariates. Chen et al. (2008) provide a careful investigation of likelihood methods for missing response and covariate data via the EM algorithm. Shardell & Miller (2008) propose a marginal modeling approach to estimate the association between a time-dependent covariate and an outcome in longitudinal studies with missing response and missing covariate, but they focus on methods with an assumption that responses are independent.

For the IPWGEE, to obtain a consistent estimate we need to correctly model the missing data process and also need to correctly model the response process given the covariates. If the missing data process model is misspecified, it can give biased estimate. While we can misspecify the distribution of the missing covariates. That means, the IPWGEE method is sensitive to the misspecification of the missing data model but robust to the misspecification of the covariate process model. For the maximum likelihood method, we do not need to specify the missing data models when missing data are MAR, but we must correctly specify the joint distribution of the response and the covariates that subject to missing. If the distribution of the covariates is misspecified, the maximum likelihood can give inconsistent estimate. That is to say, the maximum likelihood method is sensitive to the misspecification of the covariate model but robust to the misspecification of the missing data model when data are MAR.

A hybrid approach is the doubly robust estimate introduced by Lipsitz et al. (1999), in which they only considered the cross-sectional studies with a missing covariate. This is an estimating equation approach with properties similar to maximum likelihood. To obtain a consistent estimate of the regression parameters, either the missing-data model or the distribution of the missing data given the observed data must be correctly specified, which is more robust to the IPWGEE and maximum likelihood method. The literature for the doubly robust estimate includes Robins & Rotnitzky (2001), Van der Laan & Robins (2003), Scharfstein et al. (1999), Lunceford & Davidian (2004), Carpenter et al. (2006), Davidian et al. (2005), Bang & Robins (2005), and Kang & Schafer (2007), Seaman & Copas (2009). This literature, however, focuses primarily on monotone missing data patterns; Vansteelandt et al. (2007) developed regression models for the mean of repeated outcomes under nonignoreable nonmonotone nonresponse, where they focus on conducting inference about the marginal mean and the conditional mean given the baseline observed covariates. Little work has devoted to the longitudinal studies with both missing response and missing covariates. In this paper, we extend the method of Lipsitz et al. (1999) to accommodate binary longitudinal data with both missing response and missing covariates. This approach is appealing in that it can not only deal with the missing response and missing covariate problem with intermittently missing data pattern but yields the optimal estimator.

The remainder of this paper is organized as follows. In Section 2, we introduce notation and models. In Section 3, we give the forms of the estimating equations and provide details on estimation and inference. Simulation studies are given in Section 4. Data arising from an Alzheimer's diease are analyzed in the application in Section 5. Concluding remarks are made in Section 6.

2 Notation and Models

2.1 Response Process

Suppose that n individuals are to be observed, with J_i repeated measurements for subject i, $i = 1, \ldots, n$. Let $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iJ_i})'$ denote the $J_i \times 1$ binary response vector for subject i that may be missing at some time points. Let $X_i = (X_{i1}, X_{i2}, \ldots, X_{iJ_i})'$ be the covariate vector that may be missing and $Z_i = (Z'_{i1}, Z'_{i2}, \ldots, Z'_{iJ_i})'$ be the covariate matrix that are always observed, where Z_{ij} is the covariate vector for subject i at time j.

Define $\mu_{ij} = E(Y_{ij}|X_i, Z_i) = P(Y_{ij} = 1|X_i, Z_i)$, and let $\mu_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{iJ_i})'$. Provided that the mean structure of Y_{ij} depends on the covariate vector for subject i at time j (Pepe & Anderson, 1994; Robins et al., 1999), we may consider the models for the mean of the form

$$g(\mu_{ij}) = X_{ij}\beta_x + Z'_{ij}\beta_z$$

for $j = 1, \ldots, J_i, i = 1, \ldots, n$, where $\beta = (\beta_x, \beta'_z)'$ is a vector of regression parameters. Here we suppose only one covariate X_{ij} is potentially missing. Comments on how to deal with the problem when multiple covariates may be missing are given in the discussion. The variance for the response Y_{ij} is specified as

$$v_{ij} = \text{Var}(Y_{ij}|X_i, Z_i) = \mu_{ij}(1 - \mu_{ij}),$$

which depends on the regression parameter vector β.

Let $Y_{ij}^* = (Y_{ij} - \mu_{ij})/\sqrt{v_{ij}}$, $\rho_{ijk} = E(Y_{ij}^* Y_{ik}^*)$, $\rho_{ij_1j_2\cdots j_K} = E(Y_{ij_1}^* Y_{ij_2}^* \cdots Y_{ij_K}^*)$ be the Kth-order correlation among components $Y_{ij_1}, Y_{ij_2}, \ldots, Y_{ij_K}$ of Y_i, and ρ denote all the correlation parameters. For given subject i, the joint probability for a response vector Y_i can be expressed via the Bahadur representation (Bahadur 1961), which is given by

$$\begin{aligned} P(Y_i &= y_i|X_i, Z_i) = \prod_{j=1}^{J_i} \{\mu_{ij}^{y_{ij}}(1-\mu_{ij})^{1-y_{ij}}\} \cdot \{1 + \sum_{j<k} \rho_{ijk} y_{ij}^* y_{ik}^* \\ &+ \sum_{j<k<l} \rho_{ijkl} y_{ij}^* y_{ik}^* y_{il}^* + \cdots + \rho_{i1\cdots J_i} y_{i1}^* \cdots y_{iJ_i}^* \}, \end{aligned} \tag{1}$$

where y_i is a realization of Y_i, and y_{ij}^* is a realization of Y_{ij}^*. This joint density requires modeling the correlation structures of all orders. In practice, it is often the case that the second order dominates the association structure while the third and higher order association is null or nearly null. Under such circumstances, then the joint density is given by

$$P(Y_i = y_i|X_i, Z_i) = \prod_{j=1}^{J_i} \{\mu_{ij}^{y_{ij}}(1-\mu_{ij})^{1-y_{ij}}\} \cdot \{1 + \sum_{j<k} \rho_{ijk} y_{ij}^* y_{ik}^* \}. \tag{2}$$

In the following, we assume that the third and higher order association for Y_i is null, and let $C_i(\rho)$ denote the correlation matrix of Y_i.

2.2 Missing Data Process

To indicate the availability of data we let $R_{ij} = 0$ if Y_{ij} and X_{ij} are missing, $R_{ij} = 1$ if Y_{ij} is missing and X_{ij} is observed, $R_{ij} = 2$ if Y_{ij} is observed and X_{ij} is missing, and $R_{ij} = 3$ if Y_{ij} and X_{ij} are observed. Let $R_i = (R_{i1}, R_{i2}, \ldots, R_{iJ_i})'$, and $\bar{R}_{ij} = \{R_{i1}, \ldots, R_{i,j-1}\}$.

Instead of modeling the joint probability $P(R_i = r_i|Y_i, X_i, Z_i)$ for R_i directly, since we are focusing on the longitudinal setting we restrict attention to conditional models of the form $P(R_{ij} = r_{ij}|\bar{R}_{ij}, Y_i, X_i, Z_i)$ which reflect the dynamic nature of the observation process over time; we can then obtain $P(R_i = r_i|Y_i, X_i, Z_i)$ through

$$\prod_{j=2}^{J_i} P(R_{ij} = r_{ij}|\bar{R}_{ij}, Y_i, X_i, Z_i) \cdot P(R_{i1} = r_{i1}|Y_i, X_i, Z_i) \ .$$

Let $\lambda_{ijk} = P(R_{ij} = k|\bar{R}_{ij}, Y_i, X_i, Z_i)$ denote the conditional probability, $k = 0, 1, 2, 3$. We write these probabilities as conditional on the previous missing data indicators for the response and covariate, as well as the full vector of responses and covariates. The formulation thus far encompasses MCAR, MAR and MNAR mechanisms since we have written the missing data model at assessment j as depending on the full vector of responses Y_i and covariates X_i. For missing at random mechanisms we require

$$P(R_i = r_i|Y_i, X_i, Z_i) = P(R_i = r_i|Y_i^o, X_i^o, Z_i), \tag{3}$$

where Y_i^o and X_i^o represent the observed components of Y_i and X_i, respectively. However, in the longitudinal setting with our conditional formulation it is very natural to make the further assumption that

$$P(R_{ij} = r_{ij}|\bar{R}_{ij}, Y_i, X_i, Z_i) = P(R_{ij} = r_{ij}|\bar{R}_{ij}, Y_i^o, X_i^o, Z_i) \tag{4}$$

for each time point j. It can be seen that (4) implies (3), but not vice versa. Moreover, while mechanism (3) covers a larger class of MAR models than (4), models under (4) are easier to formulate and interpret. Finally, many useful models can be embedded into the class characterized by (4), and this approach has been commonly used to model missing data processes with a MAR mechanism (e.g., Robins et al., 1995). For intermittently MAR data, it is often convenient to adopt the further assumption that the missing data indicators at time j depend only on the previously observed outcomes and covariates.

To model λ_{ijk}, typically, a generalized logistic link, by using λ_{ij0} as a reference, may relate a linear function of $\bar{R}_{ij}$, Y_i, X_i and Z_i, i.e.

$$\log\left(\frac{\lambda_{ijk}}{\lambda_{ij0}}\right) = u'_{ijk}\alpha_k, \quad k = 1, 2, 3,$$

where u_{ijk} may be a subset of $\{\bar{R}_{ij}, Y_i, X_i, Z_i\}$. Let $\alpha = (\alpha'_1, \alpha'_2, \alpha'_3)'$.

Let $\pi_{ij} = P(R_{ij} = 3|Y_i, X_i, Z_i)$ be the marginal probability of observing subject i at time j, given the entire vectors of responses and covariates; it is given by

$$\pi_{ij} = \sum_{r_{i1},\ldots,r_{i,j-1}} P(R_{ij} = 3, R_{i,j-1} = r_{i,j-1}, \ldots, R_{i1} = r_{i1} | Y_i, Z_i, X_i).$$

This marginal probability can be expressed in terms of the marginal (conditional) probabilities, ${\lambda_{ijk}}'s$.

2.3 Missing Covariate Model

Since subjects can have X_i missing, we must consider the density of X_i in some situations to obtain valid analysis, where we assume the joint density of X_i given Z_i does not depend on the response vector Y_i. In practice, this joint density can be expressed as

$$P(X_i = x_i | Z_i; \gamma) = \prod_{j=2}^{J_i} P(X_{ij} = x_{ij} | \bar{X}_{ij}, Z_i; \gamma) \cdot P(X_{i1} = x_{i1} | Z_i; \gamma), \quad (5)$$

where $\bar{X}_{ij} = \{X_{i1}, \ldots, X_{i,j-1}\}$ is the history of the covariate X_{ij} until time $j - 1$, and γ is the corresponding coefficient vector.

3 Methods of Estimation

We denote the vector of all the parameters as $\theta = (\beta', \gamma', \alpha')'$. Our main interest is in estimation of β, with γ and α viewed as nuisance parameters.

3.1 Weighted Estimating Equation for the Response Parameters

Following the spirit of the IPWGEE approach of Robins, Rotnitzky, and Zhao (1995), we introduce a weight matrix $\Delta_i^*(\alpha)$ into the usual GEE to adjust for the effects of incomplete responses and covariates. That is, if we let $\Delta_i^*(\alpha) = \text{diag}(I(R_{ij} = 3)/\pi_{ij}, 1 \leq j \leq J_i)$, then the product $\Delta_i^*(Y_i - \mu_i)$ yields an adjusted contribution from subject i which involves the observed data alone.

Moreover, this element has expectation zero, and hence unbiased estimating equations for β can be obtained as

$$U^*(\beta, \alpha) = \sum_{i=1}^{n} U_i^*(\beta, \alpha) = 0, \tag{6}$$

where $U_i^*(\beta, \alpha) = D_i V_i^{-1} \Delta_i^*(\alpha)(Y_i - \mu_i)$ with $D_i = \partial \mu_i' / \partial \beta$ being a $p \times J_i$ derivative matrix, and V_i the working covariance matrix for the response Y_i.

In practice, the covariance matrix V_i is often expressed as $V_i = F_i^{1/2} C_i F_i^{1/2}$, where C_i is a working correlation matrix, and $F_i = \text{diag}(v_{ij}, j = 1, \ldots, J_i)$ and is assumed only depends on the marginal mean μ_i. When the working correlation matrix C_i is the identity matrix, (6) is computable. However, when a working independence assumption is not adopted, (6) may not be computable since elements of $D_i V_i^{-1}$ associated with the observed pairs (Y_{ij}, X_{ij}) may be unknown because they involve of other missing covariates X_{ik} $(k \neq j)$. Here we modify (6) to incorporate general working correlation matrices. We define $\Delta_i = [\delta_{ijk}]_{J_i \times J_i}$, where $\delta_{ijk} = [I(R_{ij} = 1, R_{ik} = 3) + I(R_{ij} = 3, R_{ik} = 3)]/\pi_{ijk}$ for $j \neq k$, $\delta_{ijj} = I(R_{ij} = 3)/\pi_{ij}$, and $\pi_{ijk} = P(R_{ij} = 1, R_{ik} = 3 | Y_i, X_i, Z_i) + P(R_{ij} = 3, R_{ik} = 3 | Y_i, X_i, Z_i)$. Let $M_i = F_i^{-1/2}[C_i^{-1} \bullet \Delta_i] F_i^{-1/2}$, where $A \bullet B = [a_{ij} \cdot b_{ij}]$ denotes the Hadamard product of $J_i \times J_i$ matrices $A = [a_{ij}]$ and $B = [b_{ij}]$. By introducing the condition that X_{ij} must be observed for elements in row j of $\Delta_i(\alpha)$, we ensure that all required elements of $D_i[V_i^{-1} \bullet \Delta_i(\alpha)](Y_i - \mu_i)$ can be computed.

The generalized estimating functions for β are given by

$$U(\beta, \alpha) = \sum_{i=1}^{n} U_i(\beta, \alpha) = 0 \ , \tag{7}$$

where $U_i(\beta, \alpha) = D_i M_i (Y_i - \mu_i)$. It is easy to see that estimating function (7) depends on the observed data and the parameters only, and hence is computable.

For the estimating equations (7), to obtain a consistent estimate, the missing data model needs to be correctly specified. If the missing data model is misspecified, it can yield biased estimates. Under a missing at random mechanism, Robins, Rotnitzky, and Zhao (1994, 1995), Robins and Rotnitzky (1995), Scharfstein, Rotnitzky, and Robins (1999), and Van der Laan and Robins (2003) proposed methods to improve the robustness of the inverse probability weighted estimates. The idea is to modify these inverse weighted equations by adding a tangent space of the conditional distribution of R_i, yielding an augmented estimating function which remains unbiased. With suitable choice of the appended function, we can get the doubly robust estimates. This approach has, to our knowledge, only been investigated to address the missingness with either incomplete response or covariate processes, but not both. Now, we describe ways for the double robustness for the general missingness patterns when either the covariates model or missing data model is correctly specified.

Following the same spirit of Van der Laan and Robins (2003), the general form of the augmented estimating functions for the general missingness patterns can be written as

$$\sum_{i=1}^{n}[U_i + \phi_i] = 0, \tag{8}$$

where ϕ_i is a function in the tangent space the conditional distribution of R_i with mean zero. The optimal $\phi_{i,opt}$ is chosen as the projection of U_i onto the tangent space of the conditional distribution of R_i. It is not hard to show that, in Hilbert space, $\phi_{i,opt} = E_{(Y_i^m, X_i^m | Y_i^o, X_i^o, Z_i, R_i)}[D_i N_i (Y_i - \mu_i)]$ with

$$N_i = F_i^{-1/2}[C_i^{-1} \bullet (\mathbb{1}\mathbb{1}' - \Delta_i)]F_i^{-1/2},$$

where $\mathbb{1}$ is a vector a 1's with length J_i, and Y_i^m and X_i^m denote the missing part of Y_i and X_i respectively. We then can solve estimating equations

$$S_1(\theta) = \sum_{i=1}^{n} S_{1i}(\theta) = \sum_{i=1}^{n}\{D_i M_i (Y_i - \mu_i) + E_{(Y_i^m, X_i^m | Y_i^o, X_i^o, Z_i, R_i)}[D_i N_i (Y_i - \mu_i)]\} = 0$$

to obtain the estimate of β. It can be shown that the resulting estimator for β is robust to the misspecification of either the missing data model or the covariates model. The proof if given in the Appendix.

In practice the parameters γ and α are unknown, and one must replace γ and α in with a consistent estimate. We describe how to obtain an estimate in the next subsection.

3.2 Estimation for the Nuisance Parameters

Since we are assuming the covariate is missing at random, we can obtain the estimate of γ through maximizing likelihood estimate. Note that the likelihood for subject i is $L_i(\gamma; X_i, Z_i) = P(X_i = x_i|Z_i)$. With complete data, we can solve the estimating equation $\sum_{i=1}^n \partial \log L_i(\gamma; X_i, Z_i)/\partial\gamma' = 0$ to obtain the estimate γ. With incomplete data for X_i, instead, we can solve the estimating equation

$$S_2(\gamma) = \sum_{i=1}^n S_{2i}(\gamma) = \sum_{i=1}^n [E_{\{X_i^m|X_i^o, Z_i\}} \partial \log L_i(\gamma; X_i, Z_i)/\partial\gamma'] = 0$$

to obtain the consistent estimate when the distribution for X_i is correctly specified.

For the estimation of the missing data parameter α, we can also employ the maximum likelihood estimate. Note that the log likelihood for α is given by

$$\ell(\alpha) = \sum_{i=1}^n \ell_i(\alpha) = \sum_{i=1}^n \sum_{j=1}^{J_i} \sum_{k=0}^3 I(R_{ij} = k) \log(\lambda_{ijk}),$$

and the score function is

$$S_3(\alpha) = \sum_{i=1}^n S_{3i}(\alpha) = \sum_{i=1}^n \sum_{j=1}^{J_i} \sum_{k=0}^3 \frac{I(R_{ij} = k)}{\lambda_{ijk}} \cdot \frac{\partial\lambda_{ijk}}{\partial\alpha'}.$$

Solving the estimating equation $S_3(\alpha) = 0$ leads to the maximum likelihood estimate $\hat{\alpha}$.

3.3 Estimation and Inferences

In the section we give details on the estimation and inference for the parameters. To obtain an estimate for θ, we can solve estimating equations

$$S(\hat{\theta}) = \begin{bmatrix} S_1(\hat{\theta}) \\ S_2(\hat{\gamma}) \\ S_3(\hat{\alpha}) \end{bmatrix} = \sum_{i=1}^n S_i(\theta) = \sum_{i=1}^n \begin{bmatrix} S_{1i}(\hat{\theta}) \\ S_{2i}(\hat{\gamma}) \\ S_{3i}(\hat{\alpha}) \end{bmatrix} = 0. \tag{9}$$

It can be shown that, provided the response model $p(y_i|x_i, z_i)$ is correctly specified, either the correct specification of the missing data model $p(r_i|x_i, y_i, z_i)$ or the correct specification of the covariate model $p(x_i|z_i)$ leads to the asymptotically unbiased estimate of β. The details of proof are given in the Appendix.

To solve estimating equations (9), we employ an EM algorithm (Dempster et al., 1977) if the covariate X is discrete and Montone Carlo (MC) EM (Wei and Tanner, 1990) algorithm if the covariate X is continuous. The key is that we need to calculate the conditional expectation in S_1 and S_2.

When X is discrete, then the second part in S_{1i} can be written as

$$E_{(Y_i^m, X_i^m | Y_i^o, X_i^o, Z_i, R_i)}[D_i N_i (Y_i - \mu_i)] = \sum_{(y_i^m, x_i^m)} w_{ixy} [D_i N_i (Y_i - \mu_i)]$$

where

$$\begin{aligned} w_{ixy} &= P(Y_i^m = y_i^m, X_i^m = x_i^m | Y_i^o, X_i^o, Z_i, R_i) = P(Y_i^m = y_i^m, X_i^m = x_i^m | Y_i^o, X_i^o, Z_i) \\ &= \frac{P(Y_i = y_i, X_i = x_i | Z_i)}{\sum_{(y_i^m, x_i^m)} P(Y_i = y_i, X_i = x_i | Z_i)} \\ &= \frac{P(Y_i = y_i | X_i = x_i, Z_i) P(X_i = x_i | Z_i)}{\sum_{(y_i^m, x_i^m)} [P(Y_i = y_i | X_i = x_i, Z_i) P(X_i = x_i | Z_i)]} \end{aligned}$$

can be regarded as a weight, where the distribution of $P(Y_i = y_i | X_i = x_i, Z_i)$ and $P(X_i = x_i | Z_i)$ can be obtained from (2) and (5), respectively.

We now introduce the EM algorithm to solve $S(\hat{\theta}) = 0$ as follows:

1. Obtain an initial value of the parameter $\theta = \theta^{(0)}$.

2. At the tth step, we have $\theta^{(t)}$, and calculate $w_{ixy}^{(t)} = w_{ixy}(\theta^{(t)})$ and $w_{ix}^{(t)} = w_{ix}(\theta^{(t)})$, where

$$w_{ix} = P(X_i^m = x_i^m | X_i^o, Z_i) = \frac{P(X_i = x_i | Z_i)}{\sum_{x_i^m} P(X_i = x_i | Z_i)}.$$

3. Treating $w_{ixy}^{(t)}$ and $w_{ix}^{(t)}$ as fixed, solve $S(\hat{\theta}^{(t+1)} | \hat{\theta}^{(t)}) = 0$ for $\hat{\theta}^{(t+1)}$, where

$$S(\hat{\theta}^{(t+1)} | \hat{\theta}^{(t)}) = \begin{bmatrix} S_1(\hat{\theta}^{(t+1)} | \hat{\theta}^{(t)}) \\ S_2(\hat{\theta}^{(t+1)} | \hat{\theta}^{(t)}) \\ S_3(\hat{\theta}^{(t+1)} | \hat{\theta}^{(t)}) \end{bmatrix} = \sum_{i=1}^{n} \begin{bmatrix} S_{1i}(\hat{\theta}^{(t+1)} | \hat{\theta}^{(t)}) \\ S_{2i}(\hat{\theta}^{(t+1)} | \hat{\theta}^{(t)}) \\ S_{3i}(\hat{\theta}^{(t+1)} | \hat{\theta}^{(t)}) \end{bmatrix}, \tag{10}$$

$$S_{1i}(\hat{\theta}^{(t+1)}|\hat{\theta}^{(t)}) = D_i M_i(\hat{\theta}^{(t)})(Y_i - \mu_i) + \sum_{(y_i^m, x_i^m)} w_{ixy}^{(t)}[D_i N_i(\hat{\theta}^{(t)})(Y_i - \mu_i)],$$

and

$$S_{2i}(\hat{\theta}^{(t+1)}|\hat{\theta}^{(t)}) = \sum_{x_i^m} w_{ix}^{(t)} \partial \log L_i(\gamma; x_i, z_i)/\partial\gamma'.$$

4. Iterate until convergence to, say $\hat{\theta}$, which gives the solution to $S(\hat{\theta}) = 0$.

When X is continuous, we employ the MCEM algorithm. Specifically, we solve (10) with

$$S_{1i}(\hat{\theta}^{(t+1)}|\hat{\theta}^{(t)}) = D_i M_i(\hat{\theta}^{(t)})(Y_i - \mu_i) + \int_{(y_i^m, x_i^m)} w_{ixy}^{(t)}[D_i N_i(\hat{\theta}^{(t)})(Y_i - \mu_i)] dY_i^m dX_i^m,$$

and

$$S_{2i}(\hat{\theta}^{(t+1)}|\hat{\theta}^{(t)}) = \int_{x_i^m} w_{ix}^{(t)} \partial \log L_i(\gamma; x_i, z_i)/\partial\gamma' dX_i^m,$$

where the weights become

$$\begin{aligned} w_{ixy}^{(t)} &= P(Y_i^m = y_i^m, X_i^m = x_i^m | Y_i^o, X_i^o, Z_i; \theta^{(t)}) \\ &= \frac{P(Y_i = y_i, X_i = x_i | Z_i; \theta^{(t)})}{\int_{(y_i^m, x_i^m)} P(Y_i = y_i, X_i = x_i | Z_i; \theta^{(t)}) dY_i^m dX_i^m} \\ &= \frac{P(Y_i = y_i | X_i = x_i, Z_i; \theta^{(t)}) P(X_i = x_i | Z_i; \theta^{(t)})}{\int_{(y_i^m, x_i^m)} P(Y_i = y_i | X_i = x_i, Z_i; \theta^{(t)}) P(X_i = x_i | Z_i; \theta^{(t)}) dY_i^m dX_i^m} \end{aligned}$$

and

$$w_{ix}^{(t)} = P(X_i^m = x_i^m | X_i^o, Z_i; \theta^{(t)}) = \frac{P(X_i = x_i | Z_i; \theta^{(t)})}{\int_{x_i^m} P(X_i = x_i | Z_i; \theta^{(t)}) dX_i^m}.$$

To solve (10) that equals 0, we need the integrations. In this case, rather than use numerical integration, we may employ a Monte Carlo method. To be specific, we sample (y_i^m, x_i^m) from the conditional density $w_{ixy}^{(t)}$ using the adaptive rejection algorithm of Gilks & Wild (1992). Repeat this L times, with the lth draw of (y_i^m, x_i^m) denoted by $(y_i^{ml(t)}, x_i^{ml(t)})$. Then

$$\begin{aligned} &\int_{(y_i^m, x_i^m)} w_{ixy}^{(t)}[D_i N_i(\hat{\theta}^{(t)})(Y_i - \mu_i)] dY_i^m dX_i^m \\ &= \frac{1}{L} \sum_{l=1}^{L} [D_i N_i(\hat{\theta}^{(t)})(Y_i - \mu_i)]\Big|_{(Y_i^m, X_i^m) = (y_i^{ml(t)}, x_i^{ml(t)})}, \end{aligned}$$

and

$$\int_{x_i^m} w_{ix}^{(t)} \partial \log L_i(\gamma; x_i, z_i)/\partial \gamma' dX_i^m = \frac{1}{L} \sum_{l=1}^{L} \partial \log L_i(\gamma; x_i, z_i)/\partial \gamma' \Big|_{X_i^m = x_i^{ml(t)}}.$$

To state the asymptotic properties of $\hat{\beta}$, we define

$$\Gamma(\beta, \gamma, \alpha) = E[\partial S_{1i}(\beta, \gamma, \alpha)/\partial \beta],$$

$$I_{12}(\beta, \gamma, \alpha) = E[\partial S_{1i}(\beta, \gamma, \alpha)/\partial \gamma],$$

$$I_{13}(\beta, \gamma, \alpha) = E[\partial S_{1i}(\beta, \gamma, \alpha)/\partial \alpha],$$

$$I_2(\gamma) = E[\partial S_{2i}(\gamma)/\partial \gamma],$$

$$I_3(\gamma) = E[\partial S_{3i}(\gamma)/\partial \alpha],$$

$$Q_i(\beta, \gamma, \alpha) = S_{1i}(\beta, \gamma, \alpha) - I_{12}(\beta, \gamma, \alpha) I_2^{-1}(\gamma) S_{2i}(\gamma) - I_{13}(\beta, \gamma, \alpha) I_3^{-1}(\alpha) S_{3i}(\alpha).$$

Theorem 1. Suppose that the regularity conditions state in the Appendix hold, if either the missing data model or the covariate model is correctly specified, we have

$$n^{1/2}(\hat{\beta} - \beta_0) \to N(0, \Gamma^{-1}(\beta_0, \gamma_0, \alpha_0) \Sigma [\Gamma^{-1}(\beta_0, \gamma_0, \alpha_0)]'),$$

where β_0 is the true value of β, γ_0 and α_0 are the probability limits of $\hat{\gamma}$ and $\hat{\alpha}$, and $\Sigma = E[Q_i(\beta_0, \gamma_0, \alpha_0) Q_i'(\beta_0, \gamma_0, \alpha_0)]$.

The proof is given in the Appendix. To make inferences, the matrix Γ can be consistently estimated with

$$\hat{\Gamma} = n^{-1} \sum_{i=1}^{n} [\partial S_{1i}(\hat{\theta}|\hat{\theta})/\partial \beta],$$

and Σ can be consistently estimated with $\hat{\Sigma} = n^{-1} \sum_{i=1}^{n} [\hat{Q}_i \hat{Q}_i']$, where

$$\hat{Q}_i = S_{1i}(\hat{\theta}|\hat{\theta}) - \hat{I}_{12}(\hat{\theta}|\hat{\theta}) \hat{I}_2^{-1}(\hat{\theta}|\hat{\theta}) S_{2i}(\hat{\theta}|\hat{\theta}) - \hat{I}_{13}(\hat{\theta}|\hat{\theta}) \hat{I}_3^{-1}(\hat{\alpha}) S_{3i}(\hat{\alpha}),$$

$$\hat{I}_{12}(\hat{\theta}|\hat{\theta}) = n^{-1} \sum_{i=1}^{n} [\partial S_{1i}(\hat{\theta}|\hat{\theta})/\partial \gamma],$$

$$\hat{I}_{13}(\hat{\theta}|\hat{\theta}) = n^{-1}\sum_{i=1}^{n}[\partial S_{1i}(\hat{\theta}|\hat{\theta})/\partial\alpha],$$

$$\hat{I}_{2}(\hat{\theta}|\hat{\theta}) = n^{-1}\sum_{i=1}^{n}[\partial S_{2i}(\hat{\theta}|\hat{\theta})/\partial\gamma],$$

$$\hat{I}_{3}(\hat{\alpha}) = n^{-1}\sum_{i=1}^{n}[\partial S_{3i}(\hat{\alpha})/\partial\alpha].$$

4 Numerical Studies

4.1 Performance of the Proposed Estimates

In this subsection, we evaluate the performance of the proposed method compared to other methods commonly used in practice through simulation studies. In the simulation studies, we focus on a setting where $J_i = J = 3$ and $n = 500$. We simulate the longitudinal binary responses from a model with

$$\text{logit } \mu_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 Z_{ij} \quad (11)$$

where Z_{ij} is a time variant covariate generated from $Bin(1, 0.5)$, and X_{ij} is a time variant binary covariate which may be missing at some time points and is generated from the model

$$\text{logit } \omega_{ij} = \gamma_0 + \gamma_1 X_{i,j-1} + \gamma_2 Z_{ij}, \quad (12)$$

where $\omega_{ij} = P(X_{ij} = 1|\bar{X}_{ij}, Z_{ij})$. We take $\beta_0 = \log(1.5)$, $\beta_1 = \log(0.5)$, $\beta_2 = \log(2)$, $\gamma_0 = \log(1)$, $\gamma_2 = 2$, and γ_1 varies from -2 to 2. The correlation matrix is exchangeable with correlation coefficient ρ.

For the missing data process, we take

$$\log\left(\frac{\lambda_{ijk}}{\lambda_{ij0}}\right) = \alpha_{0k} + \alpha_{1k1}I(R_{i,j-1} = 1) + \alpha_{1k2}I(R_{i,j-1} = 2) + \alpha_{1k3}I(R_{i,j-1} = 3) + \alpha_{2k}y^o_{i,j-1} + \alpha_{3k}x^o_{i,j-1}, \quad (13)$$

for $k = 1, 2, 3$, where $y^o_{i,j-1} = y_{i,j-1}$ if $y_{i,j-1}$ is observed and 0 otherwise, $x^o_{i,j-1} = x_{i,j-1}$ if $x_{i,j-1}$ is observed and 0 otherwise. The true values are taken as $\alpha_{0k} = \log(1.5)$, $\alpha_{1k1} = \log(1.5)$, $\alpha_{1k2} = \log(1.3)$, $\alpha_{1k3} = \log(1.1)$, $\alpha_{3k} = -2$, and $\alpha_{2k} = \alpha_2$ varies from -2 to 2.

In the simulations, we always assume the model for $(y_i|x_i, z_i)$ is correctly specified. We consider the following seven methods: *1)* both the missing data model and covariate model are correctly specified, which we denote $(x+, r+)$; *2)* the missing data model is correctly specified, but the model for ω is misspecified as

$$\text{logit}\ \omega_{ij} = \gamma_0^* + \gamma_2^* Z_{ij}, \tag{14}$$

which we denote $(x-, r+)$; *3)* the model for ω is correctly specified, but the missing data model is misspecified as

$$\begin{aligned}\log\left(\frac{\lambda_{ijk}}{\lambda_{ij0}}\right) &= \alpha^*_{0k} + \alpha^*_{1k1} I(R_{i,j-1} = 1) + \alpha^*_{1k2} I(R_{i,j-1} = 2) + \alpha^*_{1k3} I(R_{i,j-1} = 3) \\ &\quad + \alpha^*_{3k} x^o_{i,j-1},\end{aligned} \tag{15}$$

which we denote $(x+, r-)$; and *4)* both the model for ω is misspecified as (14), and the missing data model is misspecified as (15), which we denote $(x-, r-)$; *5)* the EM algorithm with the covariate model incorrectly specified as (14), which we denote EM$(x-)$; *6)* the simple weighted GEE (not the robust method) method that with the missing data model incorrectly specified as (15), which we denote GEE$(r-)$; *7)* the complete case analysis using maximum likelihood method, which we denote cc. In each setting, we perform 2000 simulations.

The results are reported in Tables 1 to 3, where the bias is the percent relative bias, SD is the standard deviation for the 2000 simulations, and CP represents the empirical coverage probability for 95% confidence intervals. It is seen that the $(x-, r-)$, EM$(x-)$, $GEE(r-)$ and complete case approaches yield larger biases and poor coverage probabilities; as the response association ρ increases, the performance decreases. The $(x+, r+)$, $(x+, r-)$ and $(x-, r+)$ methods provide ignorable finite sample biases, and gives good coverage probabilities; $(x-, r+)$ estimate is more efficient than $(x+, r-)$ estimate, indicating that the efficiency of the estimate is more sensitive to the misspefication of the missing data model than to the covariate model.

4.2 Impact of Model Misspecification

The validity of this algorithm depends on correct specification of the model for the response process and either the observation process and the missing covariate process. Here we investigate the impact of misspecification of the observation process model and/or the missing covariate process.

Let $\hat{\theta}^\dagger$ denote the estimator for θ when the missing data process and the covariate process are misspecified. To characterize the asymptotic bias of $\hat{\theta}^\dagger$, we use the methods of White (1982) to find the value to which $\hat{\theta}^\dagger$ converges. In the spirit of Rotnitzky and Wypij (1994), Fitzmaurice, Molenberghs, and Lipsitz (1995) and Cook, Zeng, and Yi (2004), we take the expectation of $S_i(\theta)$ with respect to the true distribution of $G = (R_i, Y_i, X_i, Z_i)$ and set it equal to zero. The solution to this equation, denoted $\theta^\dagger$, is the value to which $\hat{\theta}^\dagger$ converges in probability. If $\mathcal{G}$ is the sample space for G, and $P(g; \theta)$ is the true probability of observing the realized value g of G, then solving the equation

$$\sum_{g \in \mathcal{G}} S_i(\theta^\dagger) \cdot P(g; \theta) = 0 \tag{16}$$

gives the relationship between θ and $\theta^\dagger$, and enables one to characterize the asymptotic bias.

In this study, response measurements are featured by the same model (11) with $\rho = 0.3$; the true model for missing data indicators is (13), and the true model for the missing covariate model is (12).

Now we consider the misspecification the missing data model and the covariate model. Figures 1 and 2 plot the asymptotic percent relative biases of β_1 and β_2 against γ_1 as α_2 changes. It is seen that β_1 and β_2 are sensitive to the misspecification of the missing data and covariate models. As the absolute value of α_2 goes to 0, the relative biases decrease; as the absolute value of γ_1 goes to 0, the relative biases decrease. For fixed α_2 that are not very big, the relative biases lines are more flat, and the biases are small, indicating that the estimate is less sensitive to the misspecification of

Table 1: Empirical bias, standard deviation and coverage probabilities for seven approaches to estimation and inference with incomplete covariate and response data ($\rho = 0.6$)

			β_0			β_1			β_2		
γ_1	α_2	Method	Bias%	SD	CP%	Bias%	SD	CP%	Bias%	SD	CP%
2	2	$(x+, r+)$	1.5	0.112	95.2	0.1	0.107	94.4	-0.6	0.095	95.3
2	2	$(x+, r-)$	-1.3	0.120	94.7	-0.7	0.113	94.6	0.7	0.098	94.5
2	2	$(x-, r+)$	1.3	0.116	94.9	-0.0	0.109	94.5	-0.7	0.096	94.4
2	2	$(x-, r-)$	8.4	0.114	93.4	-5.5	0.100	93.7	-3.6	0.096	94.5
2	2	EM$(x-)$	-14.1	0.146	92.8	-51.8	0.141	84.4	28.6	0.110	85.3
2	2	GEE$(r-)$	23.7	0.196	83.8	-17.8	0.222	84.2	-2.3	0.204	94.4
2	2	cc	185.3	0.716	74.5	-42.4	0.773	85.7	12.2	0.703	94.9
2	-2	$(x+, r+)$	1.1	0.103	94.5	1.0	0.093	94.6	-0.8	0.086	94.7
2	-2	$(x+, r-)$	1.2	0.107	94.6	1.8	0.102	94.7	-1.1	0.094	94.7
2	-2	$(x-, r+)$	-1.3	0.105	94.5	-1.2	0.098	94.9	-1.4	0.088	94.6
2	-2	$(x-, r-)$	-9.5	0.108	91.4	-7.3	0.099	91.9	-4.3	0.096	94.2
2	-2	EM$(x-)$	-25.0	0.140	90.5	-73.3	0.136	80.6	26.3	0.110	88.3
2	-2	GEE$(r-)$	32.0	0.230	81.3	18.8	0.251	80.4	2.6	0.264	94.2
2	-2	cc	-272.6	0.690	89.0	72.6	0.977	96.0	28.8	0.655	95.2
-2	2	$(x+, r+)$	1.6	0.100	95.0	0.3	0.128	94.5	-1.2	0.101	95.2
-2	2	$(x+, r-)$	-1.3	0.108	94.5	-1.3	0.134	94.7	-1.3	0.110	94.6
-2	2	$(x-, r+)$	0.2	0.102	95.3	1.0	0.132	95.1	0.3	0.105	95.0
-2	2	$(x-, r-)$	7.6	0.103	94.5	5.8	0.143	94.7	-2.6	0.108	94.8
-2	2	EM$(x-)$	-31.4	0.119	88.1	-62.8	0.109	82.5	30.9	0.123	83.6
-2	2	GEE$(r-)$	51.3	0.178	78.3	-9.9	0.202	89.5	-1.3	0.200	94.4
-2	2	cc	170.0	0.506	66.7	-41.8	0.564	90.9	16.7	0.664	97.0
-2	-2	$(x+, r+)$	0.7	0.104	95.1	0.8	0.080	94.5	-0.9	0.093	94.9
-2	-2	$(x+, r-)$	-1.0	0.110	95.2	-1.6	0.088	94.9	1.6	0.102	95.0
-2	-2	$(x-, r+)$	0.4	0.105	94.4	1.0	0.084	94.8	-0.3	0.096	94.5
-2	-2	$(x-, r-)$	-20.1	0.094	91.4	12.0	0.081	92.9	3.0	0.096	93.9
-2	-2	EM$(x-)$	-41.3	0.112	84.9	-87.1	0.093	74.3	29.5	0.121	83.8
-2	-2	GEE$(r-)$	-57.5	0.235	74.4	-11.3	0.280	91.6	-4.7	0.261	93.4
-2	-2	cc	-302.0	0.876	53.8	49.9	1.077	96.8	0.4	1.218	94.6

Relative bias defined by $(\bar{\hat{\beta}} - \beta_{true})/\beta_{true} \times 100$.
SD is the standard deviation for the 2000 times simulation, which is defined by $(2000-1)^{-1}\sum_{i=1}^{2000}(\hat{\beta}^{(i)} - \bar{\hat{\beta}})^2$, where $\hat{\beta}^{(i)}$ is the ith simulation result, and $\bar{\hat{\beta}} = 2000^{-1}\sum_{i=1}^{2000}\hat{\beta}^{(i)}$.

Table 2: Empirical bias, standard deviation and coverage probabilities for seven approaches to estimation and inference with incomplete covariate and response data ($\rho = 0.3$)

			β_0			β_1			β_2		
γ_1	α_2	Method	Bias%	SD	CP%	Bias%	SD	CP%	Bias%	SD	CP%
2	2	$(x+, r+)$	1.6	0.100	95.0	0.6	0.100	94.7	-0.2	0.086	94.4
2	2	$(x+, r-)$	0.1	0.108	94.5	0.6	0.106	94.6	0.6	0.092	94.9
2	2	$(x-, r+)$	0.9	0.101	94.7	0.5	0.102	94.5	-0.4	0.088	95.3
2	2	$(x-, r-)$	6.6	0.103	94.2	-5.1	0.100	93.9	-2.5	0.084	94.6
2	2	EM$(x-)$	-10.3	0.120	91.5	-66.2	0.125	80.6	21.3	0.099	90.7
2	2	GEE$(r-)$	10.9	0.183	84.3	-8.9	0.205	84.5	2.3	0.207	94.3
2	2	cc	128.0	0.593	86.0	-19.8	0.665	92.0	17.0	0.628	96.0
2	-2	$(x+, r+)$	1.4	0.095	95.2	0.9	0.091	94.8	0.1	0.081	94.7
2	-2	$(x+, r-)$	0.8	0.109	94.6	1.0	0.101	94.7	0.6	0.083	94.8
2	-2	$(x-, r+)$	-0.6	0.097	94.7	-0.9	0.092	94.4	-0.3	0.082	94.5
2	-2	$(x-, r-)$	-7.5	0.097	93.8	-6.3	0.095	94.1	0.0	0.083	95.2
2	-2	EM$(x-)$	-10.2	0.121	91.7	-86.1	0.122	74.3	21.6	0.099	90.8
2	-2	GEE$(r-)$	15.9	0.215	81.4	11.5	0.252	86.4	-1.3	0.255	94.5
2	-2	cc	-221.1	0.955	75.8	-26.8	1.112	97.0	-27.7	1.423	93.9
-2	2	$(x+, r+)$	0.1	0.076	94.6	0.3	0.077	94.8	0.4	0.091	94.7
-2	2	$(x+, r-)$	0.8	0.077	94.4	0.1	0.080	94.6	-0.8	0.098	94.7
-2	2	$(x-, r+)$	-0.4	0.076	94.4	0.6	0.079	94.8	0.1	0.092	94.4
-2	2	$(x-, r-)$	-6.3	0.077	94.8	5.0	0.077	94.3	-1.4	0.090	94.6
-2	2	EM$(x-)$	-16.7	0.090	90.5	-65.0	0.097	80.5	26.5	0.107	88.3
-2	2	GEE$(r-)$	23.5	0.167	79.7	5.4	0.204	91.6	4.3	0.207	94.7
-2	2	cc	109.7	0.400	84.0	-37.9	0.469	94.0	5.6	0.610	91.0
-2	-2	$(x+, r+)$	0.1	0.060	95.4	0.1	0.072	95.1	0.3	0.086	94.6
-2	-2	$(x+, r-)$	0.0	0.066	94.3	0.8	0.071	94.9	0.2	0.091	94.7
-2	-2	$(x-, r+)$	1.2	0.062	94.7	0.6	0.079	94.8	-0.9	0.087	94.5
-2	-2	$(x-, r-)$	-12.4	0.076	93.4	8.4	0.077	94.1	2.0	0.087	94.2
-2	-2	EM$(x-)$	-10.6	0.088	91.4	-85.2	0.097	73.5	26.1	0.100	89.4
-2	-2	GEE$(r-)$	-34.1	0.238	74.6	-3.7	0.272	94.4	2.7	0.269	95.4
-2	-2	cc	-219.6	0.784	78.6	-27.0	1.065	97.2	0.0	0.930	94.9

Relative bias defined by $(\bar{\hat{\beta}} - \beta_{true})/\beta_{true} \times 100$.
SD is the standard deviation for the 2000 times simulation, which is defined by $(2000-1)^{-1}\sum_{i=1}^{2000}(\hat{\beta}^{(i)} - \bar{\hat{\beta}})^2$, where $\hat{\beta}^{(i)}$ is the ith simulation result, and $\bar{\hat{\beta}} = 2000^{-1}\sum_{i=1}^{2000}\hat{\beta}^{(i)}$.

Table 3: Empirical bias, standard deviation and coverage probabilities for seven approaches to estimation and inference with incomplete covariate and response data ($\rho = 0.0$)

			β_0			β_1			β_2		
γ_1	α_2	Method	Bias%	SD	CP%	Bias%	SD	CP%	Bias%	SD	CP%
2	2	$(x+, r+)$	0.3	0.091	94.8	0.4	0.099	94.5	0.2	0.082	94.5
2	2	$(x+, r-)$	0.7	0.094	94.5	1.0	0.097	94.6	1.2	0.086	94.8
2	2	$(x-, r+)$	0.5	0.090	95.2	-0.4	0.100	94.9	-0.8	0.086	95.0
2	2	$(x-, r-)$	1.3	0.090	95.4	0.4	0.097	94.4	-0.2	0.082	95.1
2	2	EM$(x-)$	28.4	0.108	91.5	-96.6	0.114	75.1	-3.0	0.081	94.6
2	2	GEE$(r-)$	2.2	0.80	93.9	-0.6	0.208	95.6	-1.0	0.206	95.3
2	2	cc	52.0	0.508	94.0	-40.8	0.599	92.0	4.5	0.565	96.0
2	-2	$(x+, r+)$	0.1	0.089	94.8	0.2	0.098	94.5	0.1	0.079	94.6
2	-2	$(x+, r-)$	-0.4	0.089	94.4	0.9	0.099	94.7	0.9	0.080	94.8
2	-2	$(x-, r+)$	-0.3	0.089	94.6	-0.2	0.098	94.2	-0.4	0.080	94.9
2	-2	$(x-, r-)$	-0.0	0.089	94.7	-0.6	0.097	94.8	-0.1	0.081	94.5
2	-2	EM$(x-)$	49.1	0.102	84.9	-109.0	0.106	71.5	-12.0	0.079	93.5
2	-2	GEE$(r-)$	3.8	0.225	95.2	3.2	0.259	94.3	2.8	0.259	94.2
2	-2	cc	-157.7	0.828	85.7	-6.8	1.021	93.3	-19.0	1.356	94.5
-2	2	$(x+, r+)$	0.3	0.060	94.9	0.5	0.078	94.3	0.2	0.090	94.5
-2	2	$(x+, r-)$	0.3	0.059	94.8	0.0	0.080	94.5	-0.3	0.091	94.7
-2	2	$(x-, r+)$	0.4	0.060	94.7	-0.1	0.079	94.5	-0.6	0.091	94.4
-2	2	$(x-, r-)$	-0.1	0.062	94.7	0.5	0.084	94.8	-0.2	0.091	94.3
-2	2	EM$(x-)$	59.3	0.080	80.2	-68.5	0.093	84.4	1.2	0.088	94.4
-2	2	GEE$(r-)$	-3.0	0.161	94.5	1.4	0.203	94.5	1.3	0.206	94.4
-2	2	cc	46.7	0.291	96.0	-58.8	0.501	93.0	-0.1	0.450	99.0
-2	-2	$(x+, r+)$	0.1	0.056	94.9	-0.3	0.077	94.5	-0.2	0.080	94.8
-2	-2	$(x+, r-)$	-0.3	0.057	94.8	-0.2	0.077	94.7	0.7	0.081	94.5
-2	-2	$(x-, r+)$	0.2	0.059	94.4	0.6	0.077	95.2	-0.3	0.080	94.2
-2	-2	$(x-, r-)$	-0.0	0.060	95.3	0.2	0.074	95.1	0.0	0.085	94.7
-2	-2	EM$(x-)$	69.4	0.081	74.7	-94.1	0.098	74.7	-8.2	0.080	93.8
-2	-2	GEE$(r-)$	1.9	0.241	94.6	1.1	0.287	94.3	2.8	0.281	94.2
-2	-2	cc	-18.1	0.820	88.6	-14.8	1.087	96.5	1.1	1.101	94.4

Relative bias defined by $(\bar{\hat{\beta}} - \beta_{true})/\beta_{true} \times 100$.

SD is the standard deviation for the 2000 times simulation, which is defined by $(2000-1)^{-1}\sum_{i=1}^{2000}(\hat{\beta}^{(i)} - \bar{\hat{\beta}})^2$, where $\hat{\beta}^{(i)}$ is the ith simulation result, and $\bar{\hat{\beta}} = 2000^{-1}\sum_{i=1}^{2000}\hat{\beta}^{(i)}$.

the covariate model; while for fixed γ_1 that are not very close to 0, the relative biases change big as α_2 changes, indicating that the estimate is more sensitive to the misspecification of the missing data model.

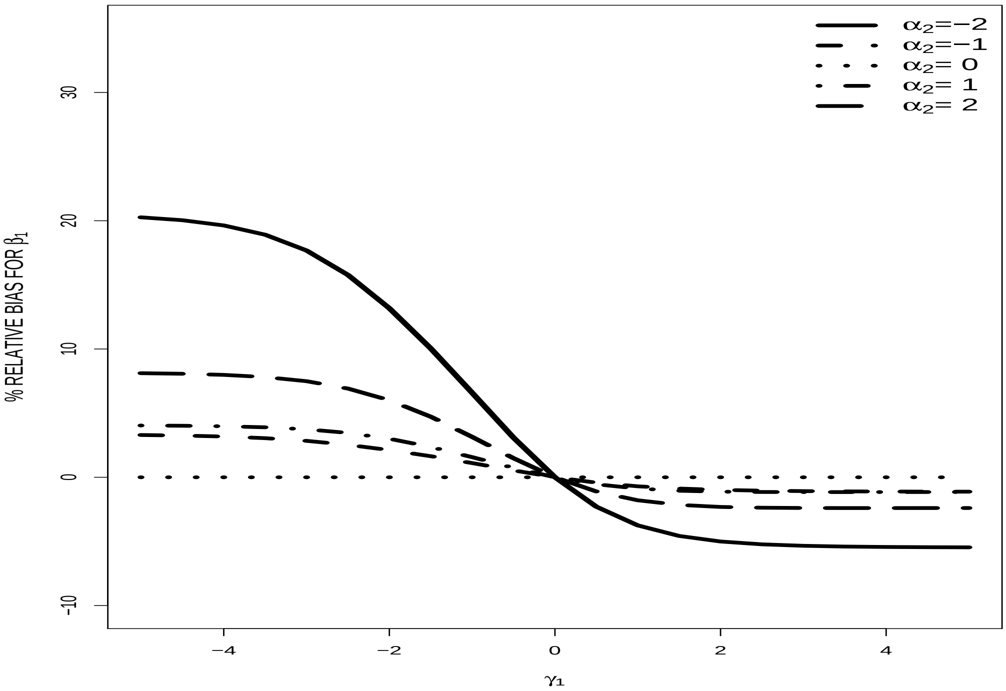

Figure 1: Asymptotic percent relative bias of β_1 with misspecified covariate model and missing data model

In summary, estimation of the response parameters is generally sensitive to misspecification of the missing data model and covariate model, although the degree of the sensitivity could be varying for different kinds of misspecification. Our asymptotic studies also suggest that if the missing data model is modeled approximately correct, then there is very good chance that the proposed method will reduce the bias with the covariate model is misspecified.

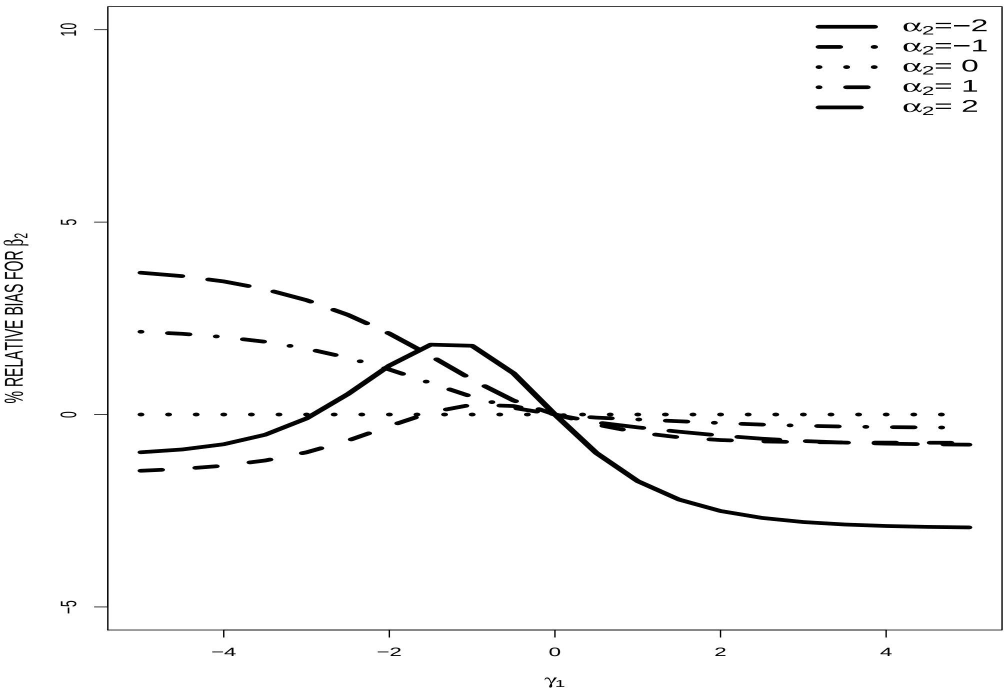

Figure 2: Asymptotic percent relative bias of β_2 with misspecified covariate model and missing data model

5 Application to an Alzheimer's Disease Study

We apply the proposed method to the National Alzheimer's Coordinating Center (NACC) Uniform Data Set (UDS). One of the goal of the study is to investigate the risk factors that influence the onset of dementia. The response is the diagnostic of dementia (Yes/No). The covariates that may influence the status of dementia include sex, congestive heart failure (CVCHF, yes/no), family history of dementia (FHDEM, yes/no), diabetes (yes/no), behavioral assessment (depression or dysphoria, yes/no), hypertension (yes/no), education (years), Mini-Mental State Exam (MMSE) score, and age. There are 16223 subjects from 29 Alzheimer's Disease Centers included at the entry of this study. Follow-up visits for subjects are scheduled at approximately one-year intervals, with up to four clinical visits at present. Due to some reasons, there are some missing data for the response and the behavioral assessment covariate. There are 8724 subjects with complete data observed. About 11.9% subjects miss both the response and behavioral assessment; about 31.2% subjects miss the response but observe behavioral assessment; about 3.2% subjects miss the behavioral assessment but observe the response; and about 53.7% subjects observe both the response and the behavioral assessment covariate.

Consider the regression model for the response process

$$\text{logit } \mu_{ij} = u_{ij}'\beta$$

where u_{ij} is the covariate vector at time point j, which include the function of sex, CVCHF, FHDEM, diabetes, depression, hypertension, education, MMSE, and age.

For the missing indicators, we build regression models

$$\log\left(\frac{\lambda_{ijk}}{\lambda_{ij0}}\right) = v_{ijk}'\alpha_k, \quad k = 1, 2, 3,$$

where v_{ijk} include function of history of the missing indicators, sex, CVCHF, FHDEM, diabetes, depression, hypertension, education, MMSE, and age.

For the covariate, we build model

$$\text{logit}\, \omega_{ij} = w_{ij}'\gamma, \quad j > 1,$$

where ω_{ij} is the conditional probability that patient i at time j is depressed given the covariate vector w_{ij} which may include function of history of the covariate, sex, CVCHF, FHDEM, diabetes, depression, hypertension, education, MMSE, and age.

In line with the simulation study, here we use three methods to analyze the data. The first method, labeled "EM", is the EM algorithm; the second method, labeled "WEE", is the doubly robust method; the third method is the complete case analysis; the results are reported in Table 4. The complete case (CC) method reveal that sex has no significant effect on the dementia, but the EM and the WEE methods reveal that it is significant; all the three methods reveal that CVCHF has no significant impact on the dementia, depression has a negative effect on the onset of dementia, MMSE has a positive effect to protect the onset of dementia, diabetes and hypertension have positive effects to protect the onset of dementia; for the family history of dementia, the CC analysis indicates that it has no significant effect, but the EM and WEE method analyses indicate that it has a negative effect on the onset of dementia; for the education level, the EM and WEE methods reveal that it has no significant effect on the onset of dementia, but the CC analysis reveals that it is not significant; for age, all three methods indicate that it has a negative effect on the onset of dementia.

For the missing data model, we carry out standard diagnostic tests for the fit of regression models by comparing a model with an expanded model to do a model selection. Here, we only list the results for the final model without reporting the tables due to the limiting space. Significance of the previous missing indicator indicates that there exists strong series dependence; sex, CVDHF, DEPD, MMSE, FHDEM, diabetes, hypertension, education, age and the observed previous response are also significant in some missing data models, indicating that data are not missing completely at random.

Table 4: Parameter estimate for the national Alzheimer's coordinating center uniform dataset: response models

	EM			WEE			CC		
Parameter	Est.	SE	p	Est.	SE	p	Est.	SE	p
(Intercept)	-0.104	0.108	0.336	-0.136	0.106	0.198	0.283	0.162	0.081
SEX(F)	-0.190	0.025	<0.001	-0.203	0.025	<0.001	-0.022	0.037	0.551
CVCHF	0.003	0.064	0.968	-0.031	0.063	0.618	-0.019	0.092	0.834
DEPRESSION	0.668	0.029	<0.001	0.679	0.029	<0.001	0.416	0.039	<0.001
MMSE	-0.005	0.001	<0.001	-0.002	0.001	<0.001	-0.021	0.001	<0.001
FHDEM	0.156	0.028	<0.001	0.181	0.028	<0.001	-0.067	0.040	0.099
DIABETE	-0.141	0.038	<0.001	-0.124	0.038	0.001	-0.168	0.054	0.002
HYPERT	-0.193	0.026	<0.001	-0.195	0.026	<0.001	-0.212	0.039	<0.001
EDUC	-0.003	0.001	0.006	-0.002	0.001	0.040	0.002	0.002	0.252
AGE	0.007	0.001	<0.001	0.006	0.001	<0.001	0.013	0.002	<0.001

6 Discussion

The consistent estimates of longitudinal data with both missing response and missing covariates under missing at random depend on the correct specification of the missing data model or the covariate model. Likelihood-based method is robust to the misspecification of the missing data process model, while the weighted estimating equation method is robust to the misspecification of the covariate model. In this paper we develop a doubly robust estimate method, which is robust to the misspeficiation of the missing data model or the misspecification of the covariate model, but not both. Simulation studies have shown that, subject to the correct specification of the response model, the estimators are consistent and empirical studies have shown that there is negligible bias in finite samples, when the missing data model is correctly specified or the covariate model is correctly specified.

The asymptotic studies have provided insight into the nature of the biases one can expect with different types of model misspecification, which suggests that there is a very good chance that our proposed method will reduce the bias with the covariate model is misspecified and the missing data model is approximately correct. Use of model diagnostics for the missing data process, perhaps most easily carried out in the MAR setting through model expansion, is warranted. It appears that empirically there is often little price to pay for introducing additional covariates into the missing data regression models. This is comforting since the more comprehensive the missing data model the more plausible it is that there is no residual dependence on the missing response, say. To provide a final check against the effects of data MNAR, sensitivity analyses can be carried out as described by Rotnitzky et al. (1998) and Scharfstein et al. (1999). It is generally not possible to check formally for the presence of a MNAR mechanism, so sensitivity analysis are required if this is a serious concern.

We focussed here primarily on estimation and inference regarding one covariate is subject to missing. Multiple covariates subject to missing are very common in practice. A future research is to extend this method to the multiple missing covariates problem. The idea is that we build missing data models to construct the weights in the weighted estimating equations, and we also need to build joint models for the covariates that are subject to missing, which is challenge in practice, especially for missing covariates with both continuous and categorical.

Appendix: Proof of the Doubly Robust Estimation Property and Theorem 1

Proof of the doubly robust estimation property.

Using the first Taylor series expansion, it can be shown that

$$n^{1/2}(\hat{\theta} - \theta) \approx n\Big\{ - \frac{\partial E[S(\theta)]}{\partial \theta} \Big\}^{-1} n^{-1/2} S(\theta),$$

which implies that

$$n^{1/2}(\hat{\beta} - \beta) \approx nI^{11} n^{-1/2} S_1(\theta) + nI^{12} n^{-1/2} S_2(\theta) + nI^{13} n^{-1/2} S_3(\alpha), \qquad (17)$$

where I^{11}, I^{12}, and I^{13} are the appropriate submatrices of $\{-\partial E[S(\theta)]/\partial\theta\}^{-1}$.

1. Missing data model is correctly specified

   Suppose missing data model and $p(y_i|x_i, z_i)$ are correctly specified, but the distribution of $(x_i|z_i)$ is not correctly specified. Then, in (9) we rewrite $E_{(Y_i^m, X_i^m|Y_i^o, X_i^o, Z_i, R_i)}[\cdot]$ and $E_{(X_i^m|X_i^o, Z_i, R_i)}[\cdot]$ as $E^*_{(Y_i^m, X_i^m|Y_i^o, X_i^o, Z_i)}[\cdot]$ and $E^*_{(X_i^m|X_i^o, Z_i)}[\cdot]$ because of the MAR assumption, where the subscript "*" represents the expectation taken over the wrongly specified distribution for $(y_i, x_i|z_i)$ and $(x_i|z_i)$.

   If the missing data model is correctly specified, we have $E[\delta_{ijk}] = 1$, and thus $E[\Delta_i] = \mathbb{1}\mathbb{1}'$, $E[M_i] = V_i^{-1}$, and $E[N_i] = 0$. So, we have $E[D_i M_i(Y_i - \mu_i)] = E[D_i V_i^{-1}(Y_i - \mu_i)] = 0$, and $E\{E^*_{(Y_i^m, X_i^m|Y_i^o, X_i^o, Z_i, R_i)}[D_i N_i(Y_i - \mu_i)]\} = E\{E^*_{(Y_i^m, X_i^m|Y_i^o, X_i^o, Z_i)}[D_i \cdot 0 \cdot (Y_i - \mu_i)]\} = 0$, if the distribution of $(y_i|x_i, z_i)$ is correctly specified. That means $E[S_1(\theta)] = 0$. It is easy to show that $E[S_3(\alpha)] = 0$.

   Now if the distribution of $(x_i|z_i)$ is incorrectly specified, then $E[S_2(\theta)] \neq 0$. However, the second term on the right hand side of (17) still has expectation equal to 0. Using the theory of partitioned matrices, it can be shown that $I^{12} = 0$ if $E[\partial S_1(\theta)/\partial\gamma] = 0$. Note that the first term of $S_1(\theta)$ does not depend on γ, so the derivative is equal to 0, hence

$$\begin{aligned} E\Big\{\frac{\partial S_{1i}(\theta)}{\partial\gamma}\Big\} &= E\Big\{\frac{\partial E^*_{(Y_i^m, X_i^m|Y_i^o, X_i^o, Z_i)} E_{(R_i|Y_i, X_i, Z_i)}[D_i N_i(Y_i - \mu_i)]}{\partial\gamma}\Big\} \\ &= E\Big\{\frac{\partial E^*_{(Y_i^m, X_i^m|Y_i^o, X_i^o, Z_i)}[0]}{\partial\gamma}\Big\} \\ &= 0. \end{aligned}$$

   Thus all the terms on the right hand side of (17) have expectation equal to 0, and $\hat{\beta}$ is asymptotically unbiased.

2. $p(x_i|z_i)$ correctly specified.

   Suppose that the distribution of $(y_i|x_i, z_i)$ and $(x_i|z_i)$ are correctly specified but the missing data model is incorrectly specified. To be specific, suppose that π_{ijk} is misspecified as π^*_{ijk}, and π_{ij}^x is misspecified as π_{ij}^{x*}, which are still functions of x_i, y_i and z_i. We define $\tilde{\Delta}_i = [\tilde{\delta}_{ijk}]$ with $\tilde{\delta}_{ijk} = [P(R_{ij} = 1, R_{ik} = 3|Y_i, X_i, Z_i) + P(R_{ij} = 3, R_{ik} = 3|Y_i, X_i, Z_i)]/\pi_{ijk}^*$ for $k \neq j$ and $\tilde{\delta}_{ijj} = P(R_{ij} = 3|Y_i, X_i, Z_i)/\pi_{ij}^*$, and

$$\tilde{\Delta}_i^* = \text{diag}(P(R_{ij} = 2 \text{ or } 3|X_i, Y_i, Z_i)/\pi_{ij}^{x*},\ j = 1, \ldots, J).$$

We show that $E[S_1(\theta)] = 0$, $E[S_2(\theta)] = 0$ and $I^{13} = 0$, implying that each term on the right hand side of (17) has 0 expectation and $\hat{\beta}$ is asymptotically unbiased.

Note that expectation of the first term of $S_{1i}(\theta)$ is

$$E[D_i M_i (Y_i - \mu_i)] = E[D_i F_i^{-1/2} [C_i \bullet \tilde{\Delta}_i] F_i^{-1/2} (y_i - \mu_i)].$$

If both $p(y_i|x_i, z_i)$ and $p(x_i|z_i)$ are correctly specified, then the joint probability $p(y_i, x_i|z_i)$ is correctly specified, and hence the expectation of the second term of $S_{1i}(\theta)$ is

$$\begin{aligned} & E\{E_{(Y_i^m, X_i^m|Y_i^o, X_i^o, Z_i, R_i)}[D_i N_i (Y_i - \mu_i)]\} \\ = \; & E\{E_{(Y_i^m, X_i^m|Y_i^o, X_i^o, Z_i)}[D_i F_i^{-1/2} [C_i \bullet (\mathbb{1}\mathbb{1}' - \tilde{\Delta}_i)] F_i^{-1/2} (Y_i - \mu_i)]\}. \end{aligned}$$

Thus we have

$$\begin{aligned} E[S_1(\theta)] &= E[D_i M_i (Y_i - \mu_i)] + E[D_i N_i (Y_i - \mu_i)] \\ &= E[D_i F_i^{-1/2} C_i^{-1} F_i^{-1/2} (Y_i - \mu_i)] \\ &= E[D_i V_i^{-1} (Y_i - \mu_i)] \\ &= 0 \end{aligned}$$

if the distribution of $(y_i|x_i, z_i)$ is correctly specified. Similarly, we can prove that $E[S_{2i}(\theta)] = 0$.

If the missing data model is misspecified, then $E[S_3(\alpha)] \neq 0$. However, the third term on the right hand side of (17) still has expectation 0 if $I^{13} = 0$. By using the theory of partitioned matrices, we can show that $I^{13} = 0$ if $E[\partial S_1(\theta)/\partial \alpha] = 0$. Note that

$$
\begin{aligned}
& E\Big\{\frac{\partial S_{1i}(\theta)}{\partial \alpha_j}\Big\} \\
= \; & E\Big\{D_i F_i^{-1/2}[C_i^{-1} \bullet \frac{\partial \Delta_i}{\partial \alpha_j}]F_i^{-1/2}(Y_i - \mu_i)\Big\} \\
& + E\Big\{E_{(Y_i^m, X_i^m | Y_i^o, X_i^o, Z_i)}[F_i^{-1/2}[C_i^{-1} \bullet \big(-\frac{\partial \Delta_i}{\partial \alpha_j}\big)]F_i^{-1/2}(Y_i - \mu_i)]\Big\} \\
= \; & 0
\end{aligned}
$$

for $j = 1, \ldots, p_3$, where $p_3 = \dim(\alpha)$. Then all the three terms on the right hand side of (17) have expectation 0, and $\hat{\beta}$ is asymptotically unbiased if the distribution of $(y_i|x_i, z_i)$ and $(x_i|z_i)$ are correctly specified.

Proof of Theorem 1.

The regularity conditions required in Theorem 1 include standard conditions that are assumed for the estimating function theory, plus the requirement for the missing data processes and covariate process. Specifically, we require $P(R_{ij} = 3|\bar{R}_{ij}, Y_i, X_i, Z_i)$ is bounded away from zero. This condition ensures that the estimating functions in (7) are bounded, which is necessary for a $\sqrt{n}$-consistent estimator. Other routine conditions are similar to those in Robins, Rotnitzky, and Zhao (1995) with a proper modification.

By standard Taylor expansion arguments we have that

$$n^{1/2}(\hat{\gamma} - \gamma_0) = -I_2^{-1}(\gamma_0) n^{-1/2} \sum_{i=1}^{n} S_{2i}(\gamma_0) + o_p(1) \tag{18}$$

and

$$n^{1/2}(\hat{\alpha} - \alpha_0) = -I_3^{-1}(\alpha_0) n^{-1/2} \sum_{i=1}^{n} S_{3i}(\alpha_0) + o_p(1). \tag{19}$$

Furthermore, based on the proof of the doubly robust properties, we have $E[S_{1i}(\beta_0, \gamma_0, \alpha_0)] = 0$ if either the missing data model or the covariate model is correctly specified. Thus, another Taylor expansion gives

$$
\begin{aligned}
0 &= n^{-1/2}\sum_{i=1}^{n} S_{1i}(\beta_0,\gamma_0,\alpha_0) + \Gamma(\beta_0,\gamma_0,\alpha_0)n^{1/2}(\hat\beta-\beta_0) + I_{12}(\beta_0,\gamma_0,\alpha_0)n^{1/2}(\hat\gamma-\gamma_0) \\
&\quad + I_{13}(\beta_0,\gamma_0,\alpha_0)n^{1/2}(\hat\alpha-\alpha_0) + o_p(1). \qquad (20)
\end{aligned}
$$

Replacing (18) and (19) into (20), we obtain

$$
\begin{aligned}
0 &= n^{-1/2}\sum_{i=1}^{n} S_{1i}(\beta_0,\gamma_0,\alpha_0) + \Gamma(\beta_0,\gamma_0,\alpha_0)n^{1/2}(\hat\beta-\beta_0) \\
&\quad - I_{12}(\beta_0,\gamma_0,\alpha_0)I_2^{-1}(\gamma_0)n^{-1/2}\sum_{i=1}^{n} S_{2i}(\gamma_0) \\
&\quad - I_{13}(\beta_0,\gamma_0,\alpha_0)I_3^{-1}(\alpha_0)n^{-1/2}\sum_{i=1}^{n} S_{3i}(\alpha_0) + o_p(1).
\end{aligned}
$$

If $\Gamma(\beta_0,\gamma_0,\alpha_0)$ is nonsingular, we have

$$
n^{1/2}(\hat\beta-\beta_0) = -\Gamma^{-1}(\beta_0,\gamma_0,\alpha_0)n^{-1/2}\sum_{i=1}^{n} Q_i(\beta_0,\gamma_0,\alpha_0) + o_p(1).
$$

Then the asymptotic distribution of $n^{1/2}(\hat\beta-\beta_0)$ follows by the Slutsky's theorem and the central limit theorem.

References

[1] Bahadur, R. R. (1961). A representation of the joint distribution of responses to n Dichotomous items. In *Studies in Item Analysis and Prediction*, H. Solomon (ed), 158–168. Stanford Mathematical Studies in the Social Sciences VI. Stanford, California: Stanford University Press.

[2] Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61**, 962–972.

[3] Carpenter, J., Kenward, M. and Vansteelandt, S. (2006). A comparison of multiple imputation and inverse probability weighting for analyses with missing data. *Journal of the Royal Statistical Society* Ser. A **169**, 571–584.

[4] Chen, Q., Ibrahim, J. G., Chen, M. and Senchaudhuri, P. (2008). Theory and inference for regression models with missing responses and covariates. *Journal of Multivariate Analysis* **99**, 1302–1331.

[5] Cook, R. J., Zeng, L. and Yi, G. Y. (2004). Marginal analysis of incomplete longitudinal binary data: a cautionary note on LOCF imputation. *Biometrics* **60**, 820–828.

[6] Davidian, M., Tsiatis, A. A. and Leon, S. (2005). Semiparametric estimation of treatment effect in a pretest-posttest study without missing data. *Statistical Science* **20**, 261–301.

[7] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society* Ser. B **39**, 1–38.

[8] Fitzmaurice, G. M., Lipsitz, S. R., Molenberghs, G. and Ibrahim J. G. (2001). Bias in estimating association parameters for longitudinal binary responses with drou-outs. *Biometrics* **57**, 15–21.

[9] Fitzmaurice, G. M., Molenberghs, G. and Lipsitz, S. R. (1995). Regression models for longitudinal binary data responses with informative drop-outs. *Journal of the Royal Statistical Society* Ser. B **57**, 691–704.

[10] Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **41**, 337–348.

[11] Horton, H. J. and Laird, N. M. (1998). Maximum likelihood analysis of generalized linear models with missing covariates. *Statistical Methods in Medical Research* **8**, 37–50.

[12] Ibrahim, J. G., Lipsitz, S. R. and Horton, N. (2001). Using auxiliary data for parameter estimation with nonignorable missing outcomes. *Applied statistics* **50**, 361–373.

[13] Kang, J. D. Y. and Schafer, J. L. (2007). Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science* **22**, 523–539.

[14] Liang, K. Y. and Zeger, S. L. (1986) Longitudinal data analysis using gerneralized linear models. *Biometrika* **73**, 13–22.

[15] Lipsitz, S. R, Ibrahim, J. G. and Zhao, L. P. (1999). A new weighted estimating equation for missing covariate data with properties similar to maximum likelihood. *Journal of the American Statistical Association* **94**, 1147–1160.

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

[17] Lunceford, J. K. and Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in Medicine* **23**, 2937–2960.

[18] Pepe, M. S. and Anderson, G. L. (1994). A cautionary note on inferecne for marginal regression models with longitudinal data and general correlated response data. *Communications in Statatistcs, Simulation and Computation* **23**, 939–951.

[19] Robins, J. M., Greenland, S. and Hu, F. C. (1999). Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome (with discussion). *Journal of the American Statistical Association* **94**, 687–712.

[20] Robins, J. M. and Rotnitzky, A. (2001). Comment on "Inference for Semiparametric Models: Some Questions and Answer," by P. J. Bickel and J. Kwon. *Statistical Sinica* **11**, 920–936.

[21] Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1995) Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90**, 106–121.

[22] Rotnitzky, A., Robins, J. M. and Scharfstein, D. O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* **93**, 1321–1339.

[23] Rotnitzky, A. and Wypij, D. (1994). A note on the bias of the estimators with missing data. *Biometrics* **50**, 1163–1170.

[24] Scharfstein, D. O., Rotnitzky, A. and Robins, J. M. (1999). Adjusting for nonignorable dropout using semiparametric nonresponse models (with discussion). *Journal of the American Statistical Association* **94**, 1096–1120.

[25] Seaman, S. and Copas, A. (2009). Doubly robust generalized estimating equations for longitudinal data. *Statistics in Medicine* **28**, 937–955.

[26] Shardell, M. and Miller, R. (2008). Weighted estimating equations for longitudinal studies with death and non-monotone missing time-dependent covariates and outcomes. *Statistics in Medicine* **27**, 1008–1025.

[27] Van der Laan, M. J. and Robins, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. Springer, New York.

[28] Vansteelandt, S. and Robins, J. M. (2007). Estimation of regression models for the mean of repeated outcomes under nonignorable nonmonotone nonresponse. *Biometrika* **94**, 841–860.

[29] Wei, G. C. G., and Tanner, M. A. (1990). A monte carlo implementation of the em algorithm and the poor mans data augmentation algorithms. *Journal of the American Statistical Association*, 85, 699–704.

[30] White, H. (1982). Maximum likelihood estimation under misspecified models. *Econometrica* **50**, 1–26.

[31] Zhao, L. P., Lipsitz, S. R. and Lew, D. (1996). Regression analysis with missing covariate data using estimating equations. *Biometrics* **52**, 1165–1182.