Published in final edited form as:

Stat Med. 2013 November 30; 32(27): . doi:10.1002/sim.5875.

A new estimation with minimum trace of asymptotic covariance matrix for incomplete longitudinal data with a surrogate process

Baojiang Chen^{a,*†} and Jing Qin^b

^aDepartment of Biostatistics, University of Nebraska Medical Center, Omaha, NE 68198, U.S.A

^bNational Institute of Allergy and Infectious Diseases, National Institute of Health, Bethesda, MD 20892, U.S.A

Abstract

Missing data is a very common problem in medical and social studies, especially when data are collected longitudinally. It is a challenging problem to utilize observed data effectively. Many papers on missing data problems can be found in statistical literature. It is well known that the inverse weighted estimation is neither efficient nor robust. On the other hand, the doubly robust (DR) method can improve the efficiency and robustness. As is known, the DR estimation requires a missing data model (i.e., a model for the probability that data are observed) and a working regression model (i.e., a model for the outcome variable given covariates and surrogate variables). Because the DR estimating function has mean zero for any parameters in the working regression model when the missing data model is correctly specified, in this paper, we derive a formula for the estimator of the parameters of the working regression model that yields the optimally efficient estimator of the marginal mean model (the parameters of interest) when the missing data model is correctly specified. Furthermore, the proposed method also inherits the DR property. Simulation studies demonstrate the greater efficiency of the proposed method compared with the standard DR method. A longitudinal dementia data set is used for illustration.

Keywords

longitudinal data; missing data; optimal; surrogate outcome

1. Introduction

Incomplete longitudinal data often arise in comparative studies because of difficulties in ascertaining outcomes at scheduled assessment times, patients refusal to undergo complete examinations, or study subjects failing to attend a scheduled clinic visit. In general, problems ensue if the mechanism leading to the missing data is dependent on the outcome. Analyses based only on individuals with complete data can lead to invalid inferences in this case. Under a *missing completely at random* mechanism [1], analyses based on generalized estimating equations (GEEs) [2] yield consistent estimates of the regression parameters. However, when the data are *missing at random* (MAR) or *missing not at random* [1], analyses based on GEE give inconsistent estimates. Robins, Rotnitzky, and Zhao [3] developed a class of inverse probability weighted GEEs (IPWGEEs), 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.

Surrogate outcomes are commonly available in longitudinal studies. In this paper, we study the problem where the outcome variable of interest is missing in a subset of the sampled data. We assume that, in addition to the collection of covariate and outcome variables, a surrogate or proxy variable for the outcome is collected for each individual. A commonly used example in dementia study can be found, where the surrogate process is the Mini-Mental State Exam (MMSE) scores but the response of interest is the dementia status. Because of various reasons, the true status of dementia may not be available for everyone; however, the MMSE score can be easily recorded. Another typical example for surrogate processes is the use of CD4 cell counts as a surrogate of survival in the study of acquired immune deficiency syndrome [4]. More discussions on surrogate variables can be found, among others, in Wittes *et al.* [5], Begg and Leung [6], Leung [7], Baker *et al.* [8], Burzykowski *et al.* [9], and Baker [10]. How to use a surrogate process to improve the robustness and estimation efficiency is a challenging problem. In statistical literature, many research papers can be found, in particularly the so-called doubly robust (DR) methods developed by Scharfstein, Rotnitzky, and Robins [11], Bang and Robins [12], Tsiatis [13], and among others. The resulting DR estimate is robust to misspecification of either the working regression model or the missing data model but not both. It is a more appealing method than the maximum likelihood method and the inverse probability weighted methods [3, 14]. The comprehensive literature for the DR estimator for the cross-sectional study, among others, includes Lunceford and Davidian [15], Carpenter, Kenward, and Vansteelandt [16], Davidian, Tsiatis, and Leon [17], and Kang and Schafer [18], and so on. For longitudinal data, Van der Laan and Robins [19] gave a general theory for the DR estimator with monotone incomplete outcome; Bang and Robins [12] showed how to perform DR inference on longitudinal data with dropout; Seaman and Copas [20] developed a DR estimating equation on longitudinal data with dropout; Chen and Zhou [21] proposed a DR estimate on longitudinal data with missing outcome and missing covariates; the other literature includes Robins and Rotnitzky [22] and Scharfstein, Rotnitzky, and Robins [11].

Despite the appealing DR property, DR estimators can exhibit severe bias when both the working regression model and the missing data model are ‘slightly’ misspecified [18]. To overcome this shortcoming, Cao, Tsiatis, and Davidian [23], Duan, Qin, and Wang [24], and Tsiatis, Davidian, and Cao [25] proposed DR estimators that improve the robustness. For a given parametric working regression model, Cao, Tsiatis, and Davidian [23] studied a DR estimator that minimizes the parameters in the working regression model such that the population mean estimation has minimum variance. Tsiatis, Davidian, and Cao [25] extended the method of [23] to monotonely coarsened data. If one is interested in estimating regression parameters instead of estimating the population mean, then the asymptotic variance derived from DR estimates is a matrix. Duan, Qin, and Wang [24] choose the minimum trace as the optimal criterion where they only considered the cross-sectional studies. In this paper, we extend the idea of minimum trace of the asymptotic covariance matrix criterion to the incomplete longitudinal data with a surrogate process. The trace of the asymptotic variance matrix is the sum of the asymptotic variances of the individual model parameters. We derive a formula for the estimator of the parameters of the working regression model that yields the optimally efficient estimator of the marginal mean model (the parameters of interest) using the minimum trace of the asymptotic covariance matrix criterion when the missing data model is correctly specified. Furthermore, this method also inhibits the DR property. Finally, the estimation procedure is very simple, which can be easily implemented.

The rest of the paper is organized as follows. In Section 2, we introduce notations and DR methods. In Section 3, we develop the optimal estimating equations in the sense of minimum trace criterion. Section 4 provides simulation studies to assess the performance the proposed method. As an illustration, we analyze the data from a dementia study in Section 5. We give the summary and concluding remarks in Section 6.

2. Notation and models

Consider a sample involving n individuals in which each individual is to be examined at J assessment times. Let Y_{ij} denote the outcome for subject i at the j th assessment, S_{ij} denote a surrogate outcome, and $\mathbf{X}_{ij}$ denote the corresponding covariate vector. For convenience, we let $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iJ})^T$, $\mathbf{S}_i = (S_{i1}, S_{i2}, \dots, S_{iJ})^T$, and $\mathbf{X}_i = (\mathbf{X}_{i1}^T, \mathbf{X}_{i2}^T, \dots, \mathbf{X}_{iJ}^T)^T$, $i = 1, \dots, n$.

The conditional mean of Y_{ij} is denoted as $\mu_{ij} = E(Y_{ij}|\mathbf{X}_{ij})$, and we let $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{iJ})^T$ denote the full vector of means. We suppose the mean of Y_{ij} depends on the covariate vector for subject i at time j through a model of the form

$$g(\mu_{ij}) = \mathbf{X}_{ij}^T \boldsymbol{\beta}, \quad (1)$$

for $j = 1, \dots, J$, $i = 1, \dots, n$, where $g(\cdot)$ is a monotone differentiable link function, $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients of interest. As commonly assumed in generalized linear model, the variance $\text{var}(Y_{ij}|\mathbf{X}_{ij})$ only depends on the marginal mean μ_{ij} . Here, it is noted that we make the implicit assumption that $E(Y_{ij}|\mathbf{X}_i) = E(Y_{ij}|\mathbf{X}_{ij})$ and $\text{var}(Y_{ij}|\mathbf{X}_i) = \text{var}(Y_{ij}|\mathbf{X}_{ij})$ as is commonly carried out [26].

Similarly, one can model the conditional expectation of Y_{ij} given S_{ij} and $\mathbf{X}_{ij}$ as

$$E(Y_{ij}|S_{ij}, \mathbf{X}_{ij}) = \varphi(S_{ij}, \mathbf{X}_{ij}; \boldsymbol{\gamma}),$$

where φ is a given function and $\boldsymbol{\gamma}$ is an unknown parameter of dimension q . Here, we also make the assumption that the conditional expectation of Y_{ij} given $\mathbf{S}_i$ and $\mathbf{X}_i$ depends only on the surrogate outcome S_{ij} and covariate $\mathbf{X}_{ij}$. Let $\mathbf{R}_{ij} = (R_{ij1}, \dots, R_{ijJ})^T$, where $R_{ijj} = (S_{ij}, \mathbf{X}_{ij}; \boldsymbol{\gamma})$.

One challenge for this kind of study is the incomplete data problem. Here, we only consider the incomplete outcome problem, and we always observe the covariates and surrogate process. We will discuss missing outcome, missing surrogate process, or missing covariates problem in Section 6. To indicate the availability of data, we let $R_{ij} = 1$ if Y_{ij} is observed, and 0 otherwise. Here, we consider a general missing pattern (intermittent or monotone) for the outcome. We consider a special case of the MAR mechanism, where we assume

$$P(R_{ij}=1|\bar{\mathbf{R}}_{i,j-1}, \mathbf{S}_i, \mathbf{Y}_i, \mathbf{X}_i) = P(R_{ij}=1|\bar{\mathbf{R}}_{i,j-1}, \mathbf{S}_i, \mathbf{X}_i), \quad (2)$$

with $\mathbf{R}_{ij} = \{R_{ij1}, \dots, R_{ijJ}\}$. Let $R_{ij} = P(R_{ij}=1|\bar{\mathbf{R}}_{i,j-1}, \mathbf{S}_i, \mathbf{X}_i)$; we then can build a generalized linear model for R_{ij} say, logistic regression, as follows:

$$\text{logit } \lambda_{ij} = \mathbf{u}_{ij}^T \boldsymbol{\alpha}, \quad (3)$$

where $\mathbf{u}_{ij}$ includes $\{\mathbf{S}_i, \mathbf{R}_{ij-1}, \mathbf{X}_i\}$ and $\boldsymbol{\beta}$ is an unknown vector of parameters. Define $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{ij})^T$. Let $p_{ij} = P(R_{ij} = 1 | \mathbf{S}_i, \mathbf{X}_i)$ be the marginal probability that subject i at time j is observed, which can be expressed by p_{ij} 's.

2.1. Inverse probability weighted generalized estimating equations

With complete data, Liang and Zeger [2] developed the GEE method for the estimation of $\boldsymbol{\beta}$, which is to solve

$$\sum_{i=1}^n \mathbf{D}_i \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) = 0, \quad (4)$$

where $\mathbf{D}_i = \partial \boldsymbol{\mu}_i^T / \partial \boldsymbol{\beta}$ and $\mathbf{V}_i$ is a working covariance matrix of $\mathbf{Y}_i$. In general, the working covariance matrix can be expressed as

$$\mathbf{V}_i = c(\xi) \mathbf{F}_i^{1/2} \mathbf{G}_i(\rho) \mathbf{F}_i^{1/2},$$

where $c(\cdot)$ is a known function, ξ is a scaled parameter, $\mathbf{F}_i$ is a $J \times J$ diagonal matrix with elements $v_{ij} = \text{Var}(Y_{ij})$, and $\mathbf{G}_i(\cdot)$ is a $J \times J$ working correlation matrix that is fully specified by the vector of parameters ρ .

Liang and Zeger [2] showed that solving (4) will yield a consistent estimator for $\boldsymbol{\beta}$ when the marginal mean model is correctly specified no matter the working correlation matrix is misspecified or not, and under a missing completely at random mechanism [1], in general, complete case analyses based on GEE [2] yield consistent estimates of the regression parameters. However, when the data are MAR [1], in general, analyses based on GEE give inconsistent estimates. To overcome this problem, Robins, Rotnitzky, and Zhao [3] developed the IPWGEE approach, in which a weight matrix $\boldsymbol{\Delta}_i = \text{diag}(R_{ij}/p_{ij}, j = 1, \dots, J)$ is introduced into the usual GEE (4) to adjust for effects of incomplete outcomes; then, the product $\boldsymbol{\Delta}_i (\mathbf{Y}_i - \boldsymbol{\mu}_i)$ 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 $\boldsymbol{\beta}$ can be obtained as

$$\sum_{i=1}^n \mathbf{D}_i \mathbf{V}_i^{-1} \boldsymbol{\Delta}_i (\mathbf{Y}_i - \boldsymbol{\mu}_i) = 0. \quad (5)$$

To obtain a consistent estimator of $\boldsymbol{\beta}$ by solving (5), the missing data model must be correctly specified. If the missing data model is misspecified, in general, the resultant estimator is inconsistent.

2.2. Doubly robust estimates

Under a MAR mechanism, Robins, Rotnitzky, and Zhao [3], Scharfstein, Rotnitzky, and Robins [11], and Van der Laan and Robins [19] proposed methods to improve the robustness of the inverse probability weighted estimators. The idea is to modify the IPWGEE by adding an appended function in the tangent space of the conditional distribution of R_{ij} , yielding an augmented estimating function, which remains unbiased. With suitable choice of the appended function, we can obtain the DR estimators. Specifically, following the same spirit of Van der Laan and Robins [19], to obtain a DR estimator, we solve for the following augmented estimating functions:

$$\sum_{i=1}^n \mathbf{D}_i \mathbf{V}_i^{-1} [\Delta_i (\mathbf{Y}_i - \mu_i) + (\mathbf{I} - \Delta_i) (\varphi_i(\mathbf{S}_i, \mathbf{X}_i; \gamma) - \mu_i(\mathbf{X}_i; \beta))] = 0, \quad (6)$$

where $\mathbf{I}$ is the identity matrix of dimension J . Solving (6) for γ yields a DR estimate $\hat{\gamma}$. That is, solving the estimating equation (6) will produce a consistent estimator of γ if either the missing data model or the working regression model is correctly specified [19].

To solve for γ from the augmented estimating equation (6), with the parameter γ replaced by the maximum-likelihood estimation $\hat{\gamma}$ in the weight matrix $\mathbf{V}_i(\gamma)$, one must firstly estimate in the working regression model β . Note that the estimating equation (6) has mean 0 for any choice of β if the missing data model (3) is correctly specified. A typical method is to use the ordinary least square (OLS) estimate with the available data, which is equivalent to solving the following estimating equation

$$\sum_{i=1}^n \frac{\partial \varphi_i^T}{\partial \gamma} \Delta_{1i} (\mathbf{Y}_i - \varphi_i) = 0, \quad (7)$$

where $\Delta_{1j} = \text{diag}(R_{1j}, j = 1, \dots, J)$. Following Cao, Tsiatis, and Davidian [23] and Duan, Qin, and Wang [24], we can easily argue that the solution of equation (6) with γ replaced by the estimate produced from (7) is consistent if the missing data model is correctly specified but not optimal when β is incorrectly specified. In the following, we will develop optimal estimating equations for γ in the sense of having minimum trace of asymptotic covariance matrix of $\hat{\gamma}$.

3. Optimal estimating equations

We first consider the case that the missing probability function is fully known; then, we relax it to the case that the missing probability function is known up to an estimated parameter.

3.1. Missing probability function is fully known

Here, we provide a derivation of the optimal estimating equation when the missing probability is fully known.

Let $\hat{\gamma}$ be an estimate of γ in the interior compact set of γ . Plugging $\hat{\gamma}$ into (6) and holding fixed, by usual expansion with respect to γ around the true value γ_0 , we have

$$\begin{aligned} 0 &= n^{-1/2} \sum_{i=1}^n \mathbf{D}_i \mathbf{V}_i^{-1} [\Delta_i (\mathbf{Y}_i - \mu_i(\mathbf{X}_i; \hat{\beta})) + (\mathbf{I} - \Delta_i) (\varphi_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma}) - \mu_i(\mathbf{X}_i; \hat{\beta}))] \\ &= n^{-1/2} \sum_{i=1}^n \mathbf{D}_i \mathbf{V}_i^{-1} [\Delta_i (\mathbf{Y}_i - \mu_i(\mathbf{X}_i; \beta_0)) + (\mathbf{I} - \Delta_i) (\varphi_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma}) - \mu_i(\mathbf{X}_i; \beta_0))] \\ &\quad - \left\{ n^{-1} \sum_{i=1}^n \mathbf{D}_i \mathbf{V}_i^{-1} \mathbf{D}_i^T \right\} \left\{ n^{1/2} (\hat{\beta} - \beta_0) \right\} + o_p(1). \end{aligned}$$

Under the regularity conditions in the Appendix of Robins, Rotnitzky, and Zhao [3], we have

$$n^{-1} \sum_{i=1}^n \mathbf{D}_i \mathbf{V}_i^{-1} \mathbf{D}_i^T \xrightarrow{p} E_X \{ \mathbf{D} \mathbf{V}^{-1} \mathbf{D} \} \equiv \sum_0.$$

Thus, we have

$$n^{1/2}(\hat{\beta} - \beta_0) = n^{-1/2} \sum_{i=1}^n \sum_0^{-1} \mathbf{W}_i + o_p(1), \quad (8)$$

where

$$\mathbf{W}_i = \mathbf{D}_i \mathbf{V}_i^{-1} [\Delta_i (\mathbf{Y}_i - \mu_i(\mathbf{X}_i; \beta_0)) + (\mathbf{I} - \Delta_i) (\varphi_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma}) - \mu_i(\mathbf{X}_i; \beta_0))].$$

In the following derivation, for simplicity, we ignore the subscript i if there is no confusion. Note that $\text{cov}_{(\mathbf{R}, \mathbf{Y}, \mathbf{S}, \mathbf{X})}(\mathbf{W}) = E_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} \{ \text{cov}_{(\mathbf{R}|\mathbf{Y}, \mathbf{S}, \mathbf{X})}(\mathbf{W}) \} + \text{cov}_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} \{ E_{(\mathbf{R}|\mathbf{Y}, \mathbf{S}, \mathbf{X})}(\mathbf{W}) \} = \mathbf{W}_1 + \mathbf{W}_2$, with

$$\begin{aligned} \tilde{\mathbf{W}}_1 &= E_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} \{ \text{cov}_{(\mathbf{R}|\mathbf{Y}, \mathbf{S}, \mathbf{X})}(\mathbf{W}) \} \\ &= E_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} \left\{ \text{cov}_{(\mathbf{R}|\mathbf{Y}, \mathbf{S}, \mathbf{X})} \left(\mathbf{D} \mathbf{V}^{-1} \Delta (\mathbf{Y} - \varphi(\mathbf{S}, \mathbf{X}; \hat{\gamma})) \right) \right\} \\ &= E_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} \{ \text{cov}_{(\mathbf{R}|\mathbf{Y}, \mathbf{S}, \mathbf{X})}(\mathbf{D} \mathbf{V}^{-1} \Psi \mathbf{d}) \} \\ &= E_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} \{ \mathbf{D} \mathbf{V}^{-1} \Psi \sum_1 \Psi \mathbf{V}^{-1} \mathbf{D}^T \}, \end{aligned}$$

where $\sum_0 = \text{diag}(\mathbf{Y} - \varphi(\mathbf{S}, \mathbf{X}; \hat{\gamma}))$, $\mathbf{d} = (R_1/\sum_1, \dots, R_J/\sum_J)^T$, and $\sum_1 = \text{cov}(\mathbf{d})$; and

$$\begin{aligned} \tilde{\mathbf{W}}_2 &= \text{cov}_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} \{ E_{(\mathbf{R}|\mathbf{Y}, \mathbf{S}, \mathbf{X})}(\mathbf{W}) \} \\ &= \text{cov}_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} \{ (\mathbf{D} \mathbf{V}^{-1} (\mathbf{Y} - \mu(\mathbf{X}; \beta_0))) \}. \end{aligned}$$

We now consider the minimum trace criterion when deriving optimal estimators. It follows from (8) that the asymptotic covariance matrix of $n^{1/2}(\hat{\beta} - \beta_0)$ is

$$E_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} \left\{ \sum_0^{-1} \mathbf{D} \mathbf{V}^{-1} \Psi \sum_1 \Psi \mathbf{V}^{-1} \mathbf{D}^T \sum_0^{-1T} \right\} + \sum_0^{-1} \text{cov}_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} \{ (\mathbf{D} \mathbf{V}^{-1} (\mathbf{Y} - \mu(\mathbf{X}; \beta_0))) \} \sum_0^{-1T}. \quad (9)$$

Clearly, only the first term of (9) involves $\sum_1$. For this reason, we now focus on the first term, of which the trace is

$$\text{tr} \left[E_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} \left\{ \sum_0^{-1} \mathbf{D} \mathbf{V}^{-1} \Psi \sum_1 \Psi \mathbf{V}^{-1} \mathbf{D}^T \sum_0^{-1T} \right\} \right].$$

To minimize this trace with respect to β_k , $k = 1, \dots, q = \dim(\dots)$, we take the first derivative of the previous trace and set it to 0. Specifically, by noting that $d \text{tr}(x) = \text{tr}(dx)$, we have,

$$\begin{aligned}
0 &= \partial \text{tr} \left[E_{(Y, S, X)} \left\{ \sum_0^{-1} \mathbf{D} \mathbf{V}^{-1} \boldsymbol{\Psi} \sum_1 \boldsymbol{\Psi} \mathbf{V}^{-1} \mathbf{D}^T \sum_0^{-1T} \right\} \right] / \partial \gamma_k \\
&= \text{tr} \left[E_{(Y, S, X)} \left\{ 2 \sum_0^{-1} \mathbf{D} \mathbf{V}^{-1} \boldsymbol{\Psi}_k \sum_1 \boldsymbol{\Psi} \mathbf{V}^{-1} \mathbf{D}^T \sum_0^{-1T} \right\} \right] \\
&= 2 \text{tr} \left[E_{(Y, S, X)} \left\{ \mathbf{V}^{-1} \mathbf{D}^T \sum_0^{-1T} \sum_0^{-1} \mathbf{D} \mathbf{V}^{-1} \boldsymbol{\Psi}_k \sum_1 \boldsymbol{\Psi} \right\} \right] \\
&= 2 E_{(Y, S, X)} \left[\mathbf{A}_k^T (\mathbf{Y} - \varphi(\mathbf{S}, \mathbf{X}; \gamma)) \right],
\end{aligned}$$

where $\mathbf{A}_k = \text{diag}(-(\mathbf{S}, \mathbf{X}; \gamma) / \mathbf{A}_{kj}, j = 1, \dots, J)$, $\mathbf{A}_k = (a_{k1}, \dots, a_{kJ})^T$, and a_{kj} is the (j, j) th element of the matrix $\left(\mathbf{V}^{-1} \mathbf{D}^T \sum_0^{-1T} \sum_0^{-1} \mathbf{D} \mathbf{V}^{-1} \boldsymbol{\Psi}_k \sum_1 \right), j = 1, \dots, J$.

Let $\mathbf{A}_i = (\mathbf{A}_{i1}^T, \dots, \mathbf{A}_{iJ}^T)^T$, where $\mathbf{A}_{ik}$ is a realization of $\mathbf{A}_k$ for subject i . Because the optimal estimating equation works only for complete data, here, we consider a modified optimal estimating equation for

$$\sum_{i=1}^n \mathbf{A}_i \Delta_i (\mathbf{Y}_i - \varphi_i(\mathbf{S}_i, \mathbf{X}_i; \gamma)) = 0, \quad (10)$$

by plugging the weight matrix $\hat{w}_i$ into the complete data optimal estimating equation. Estimation of the parameter γ by solving (10) can be carried out using the Newton–Raphson algorithm through iteration steps, which is easy to implement in practice. In each iteration step, we replace w_i by $\hat{w}_i$ and $\mathbf{A}_i$ by $\hat{\mathbf{A}}_i$ where $\hat{w}_i$ is w_i with $\hat{w}_i$ plugged in,

$\hat{\mathbf{A}}_i = (\hat{\mathbf{A}}_{i1}^T, \dots, \hat{\mathbf{A}}_{iJ}^T)^T$, $\hat{\mathbf{A}}_{ik} = (\hat{a}_{ik1}, \dots, \hat{a}_{ikJ})^T$, $\hat{a}_{ikj}$ is the (j, j) th element of the matrix $\left(\hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}_i^T \sum_0^{-1T} \sum_0^{-1} \hat{\mathbf{D}}_i \hat{\mathbf{V}}_i^{-1} \hat{\boldsymbol{\Psi}}_{ik} \sum_1 \right), j = 1, \dots, J$, $\sum_1 = n^{-1} \sum_{i=1}^n (\mathbf{d}_i - 1)(\mathbf{d}_i - 1)^T$, $\sum_0 = n^{-1} \sum_{i=1}^n [\hat{\mathbf{D}}_i \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}_i]$, $\mathbf{D}_i$ and $\mathbf{V}_i$ are $\mathbf{D}_i$ and $\mathbf{V}_i$ with $\hat{w}_i$ plugged in, respectively, and $\hat{\boldsymbol{\Psi}}_{ik} = \text{diag}(-(\hat{S}_{ij}, \hat{X}_{ij}; \gamma) / \hat{\mathbf{A}}_{ikj}, j = 1, \dots, J)$. When the missing data model is correctly specified, the modified estimating equation estimate is optimal. When the missing data model is incorrectly specified, in simulations, the modified optimal estimating equation estimate is demonstrated to be better than the OLS estimate, even if it may not actually be optimal.

3.2. Missing probability function is known up to a parameter

When the missing probability function is known up to a parameter γ , we have to consider the impact of estimating the extra nuisance parameter γ on the estimation of β . In the following, we will derive an optimal estimating equation when the missing data parameter is estimated from model (3).

Assume that the missing data model is correctly specified. Typically, estimation for the missing data parameter γ can be carried out by maximum likelihood method, which is equivalent to solving the score function $\sum_{i=1}^n \mathbf{U}_{2i}(\alpha) = 0$, with $\mathbf{U}_{2i}(\alpha) = \mathbf{v}_i(\mathbf{R}_i - \eta_i)$, where $\mathbf{v}_i$ is a function of η_i . The Fisher information matrix for α can be expressed by $\mathbf{I}(\alpha) = E[\mathbf{U}_{2i}(\alpha) \mathbf{U}_{2i}^T(\alpha)]$.

In the following, we provide a sketch derivation of the optimal estimating equation when the missing data parameter is estimated from model (3). For detailed derivation, please see the Appendix.

Subject to regularity conditions in the Appendix, and by usual expansion of (6) with respect to (β, γ) around the true value (β_0, γ_0) , we have

$$\begin{aligned} 0 = & n^{-1/2} \sum_{i=1}^n \mathbf{D}_i \mathbf{V}_i^{-1} [\Delta_i(\alpha_0)(\mathbf{Y}_i - \mu_i(\mathbf{X}_i; \beta_0)) + (\mathbf{I} - \Delta_i(\alpha_0))(\varphi_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma}) - \mu_i(\mathbf{X}_i; \beta_0))] \\ & - \sum_0 \{n^{1/2}(\hat{\beta} - \beta_0)\} \\ & + \Gamma n^{1/2}(\hat{\alpha} - \alpha_0) + o_p(1), \end{aligned} \quad (11)$$

where $\alpha = (\alpha_1, \dots, \alpha_q)^T$, $q = \dim(\alpha)$, with

$$\Gamma_k = E_{(\mathbf{R}, \mathbf{Y}, \mathbf{S}, \mathbf{X})} [\mathbf{D}_i \mathbf{V}_i^{-1} \Delta_{ik}(\mathbf{Y}_i - \varphi_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma}))],$$

and $\Delta_{ik} = \Delta_i(\alpha_k)$, $k = 1, \dots, q$.

By standard Taylor expansion arguments for the missing data score function, we have that

$$n^{1/2}(\hat{\alpha} - \alpha_0) = \mathbf{I}^{-1}(\alpha_0) n^{-1/2} \sum_{i=1}^n \mathbf{U}_{2i}(\alpha_0) + o_p(1). \quad (12)$$

Replacing (12) into (11), and after some algebra, we have

$$n^{1/2}(\hat{\beta} - \beta_0) = n^{-1/2} \sum_{i=1}^n \sum_0^{-1} \mathbf{W}_i + o_p(1),$$

where

$$\mathbf{W}_i = \mathbf{D}_i \mathbf{V}_i^{-1} [\Delta_i(\alpha_0)(\mathbf{Y}_i - \mu_i(\mathbf{X}_i; \beta_0)) + (\mathbf{I} - \Delta_i(\alpha_0))(\varphi_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma}) - \mu_i(\mathbf{X}_i; \beta_0))] + \Gamma \mathbf{I}^{-1}(\alpha_0) \mathbf{U}_{2i}(\alpha_0).$$

Now we derive the optimal estimating equation for β . In the same spirit of derivation when the missing data parameter is known, we consider to take the variance of $\mathbf{W}_i$ with respect to $\mathbf{R}$ given $(\mathbf{Y}, \mathbf{S}, \mathbf{X})$. It is noted that

$$\begin{aligned} \mathbf{W}_i &= \mathbf{D}_i \mathbf{V}_i^{-1} \Delta_i(\alpha_0)(\mathbf{Y}_i - \varphi_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma})) + \mathbf{D}_i \mathbf{V}_i^{-1} (\varphi_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma}) - \mu_i(\mathbf{X}_i; \beta_0)) + \Gamma \mathbf{I}^{-1}(\alpha_0) \mathbf{v}_i(\mathbf{R}_i - \lambda_i) \\ &= \mathbf{D}_i \mathbf{V}_i^{-1} \Psi_i \mathbf{d}_i + \Gamma \mathbf{I}^{-1}(\alpha_0) \mathbf{v}_i \Delta_{2i} \mathbf{d}_i + \mathbf{D}_i \mathbf{V}_i^{-1} (\varphi_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma}) - \mu_i(\mathbf{X}_i; \beta_0)) - \Gamma \mathbf{I}^{-1}(\alpha_0) \mathbf{v}_i \lambda_i \\ &= \mathbf{H}_i \mathbf{d}_i + \mathbf{D}_i \mathbf{V}_i^{-1} (\varphi_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma}) - \mu_i(\mathbf{X}_i; \beta_0)) - \Gamma \mathbf{I}^{-1}(\alpha_0) \mathbf{v}_i \lambda_i, \end{aligned}$$

where $\Delta_{2i} = \text{diag}(\Delta_{2i}, \dots, \Delta_{2i})$, $j = 1, \dots, J$, and $\mathbf{H}_i = \mathbf{D}_i \mathbf{V}_i^{-1} \Psi_i + \Gamma \mathbf{I}^{-1}(\alpha_0) \mathbf{v}_i \Delta_{2i}$.

In the following derivation, for simplicity, we ignore the subscript i if there is no confusion. Taking the variance of $\mathbf{W}$ with respect to $\mathbf{R}$ given $(\mathbf{Y}, \mathbf{S}, \mathbf{X})$, we can show that the part of the covariance of $\mathbf{W}$, which involves the nuisance parameter γ , is $E_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})}[\mathbf{H}^{-1} \mathbf{H}^T]$. Now, we

take the first derivative of the trace of the asymptotic covariance matrix of $n^{1/2}(\hat{\gamma} - \gamma_0)$, with respect to γ_k , $k = 1, \dots, q$, which is

$$\partial \text{tr} \left[\sum_0^{-1} \text{COV}_{(R, Y, S, X)}(\mathbf{W}) \sum_0^{-1T} \right] / \partial \gamma_k = 2E_{(Y, S, X)} [\mathbf{A}_k (\mathbf{Y} - \boldsymbol{\varphi}(\mathbf{S}, \mathbf{X}; \gamma))] + 2\mathbf{B}_k,$$

where $\mathbf{B}_k = \text{tr} E_{(Y, S, X)} \left[\sum_0^{-1} \mathbf{D} \mathbf{V}^{-1} \boldsymbol{\Psi}_k \sum_1 \left(\boldsymbol{\Delta}_{2i} \mathbf{v}_i^T \mathbf{I}^{-1}(\alpha_0) \mathbf{I}^T \right) \sum_0^{-1T} \right]$. Let $\mathbf{B} = (\mathbf{B}_1, \dots, \mathbf{B}_q)^T$; then, we have

$$\partial \text{tr} \left[\sum_0^{-1} \text{COV}_{(R, Y, S, X)}(\mathbf{W}) \sum_0^{-1T} \right] / \partial \gamma = 2E_{(Y, S, X)} [\mathbf{A}(\mathbf{Y} - \boldsymbol{\varphi}(\mathbf{S}, \mathbf{X}; \gamma)) + 2\mathbf{B}].$$

With missing data, we consider a modified optimal estimating equation for γ as follows:

$$\mathbf{0} = \sum_{i=1}^n [\mathbf{A}_i \boldsymbol{\Delta}_i (\mathbf{Y}_i - \boldsymbol{\varphi}_i(\mathbf{S}_i, \mathbf{X}_i; \gamma)) + \boldsymbol{\Delta}_i \mathbf{B}_i], \quad (13)$$

where $\mathbf{A}_i$ is same as in Section 3.1, $\mathbf{B}_i = (B_{i1}, \dots, B_{iq})^T$, with

$B_{ik} = \text{tr} \left[\sum_0^{-1} \mathbf{D}_i \mathbf{V}_i^{-1} \boldsymbol{\Psi}_{ik} \sum_1 \left(\boldsymbol{\Delta}_{2i} \mathbf{v}_i^T \mathbf{I}^{-1}(\alpha_0) \mathbf{I}^T \right) \sum_0^{-1T} \right]$, $k = 1, \dots, q$. Similarly, estimation of the parameter γ by solving (13) can be carried out using the Newton–Raphson algorithm through iteration steps. In each iteration step, for a given value $\hat{\gamma}_j$ and the maximum likelihood estimate $\hat{\gamma}$, we replace $\mathbf{A}_i$ by $\hat{\mathbf{A}}_i$, $\boldsymbol{\Delta}_i$ by $\hat{\boldsymbol{\Delta}}_i$ and $\mathbf{B}_i$ by $\hat{\mathbf{B}}_i$ where $\hat{\mathbf{B}}_i = (B_{i1}, \dots, B_{iq})^T$, $\hat{B}_{ik} = \text{tr} \left[\widehat{\mathbf{D}}_i \widehat{\mathbf{V}}_i^{-1} \widehat{\boldsymbol{\Psi}}_{ik} \sum_1 \left(\widehat{\boldsymbol{\Delta}}_{2i} \widehat{\mathbf{v}}_i^T \mathbf{I}^{-1}(\hat{\alpha}) \mathbf{I}^T \right) \widehat{\sum}_0^{-1T} \right]$, $\widehat{\boldsymbol{\Delta}}_{2i} = \text{diag}(\hat{\gamma}_{ij})$, $j = 1, \dots, J$, $\widehat{\mathbf{v}}_i = (\hat{v}_{i1}, \dots, \hat{v}_{iq})^T$ with $\hat{\mathbf{v}}_i = n^{-1} \sum_{i=1}^n [\widehat{\mathbf{D}}_i \widehat{\mathbf{V}}_i^{-1} \widehat{\boldsymbol{\Delta}}_{ik} (\mathbf{Y}_i - \boldsymbol{\varphi}_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma}))]$, and $\widehat{\mathbf{v}}_i$ and $\widehat{\boldsymbol{\Delta}}_{ik}$ are $\mathbf{v}_i$ and $\boldsymbol{\Delta}_{ik}$ with γ plugged in.

Specifically, we may employ a two-stage estimation procedure as follows:

- a. In stage 1, we estimate for the unknown missing data parameter γ using the maximum likelihood method, denote the estimate $\hat{\gamma}$;
- b. In stage 2, we solve for parameters γ and γ through the following iteration steps:
 - (b1) For given $\hat{\gamma}$ and $\hat{\gamma}$, solve estimating equation (13) for γ ;
 - (b2) Substitute $\hat{\gamma}$ and $\hat{\gamma}$ into (6) for γ ;
 - (b3) Iterate (b1) and (b2) until convergence.

We have the following asymptotic results.

Theorem 1—Suppose that the regularity conditions stated in the Appendix hold, if either the missing data model or the working regression model is correctly specified, and if the working regression parameter γ is known to be γ_0 , we have

$$n^{1/2}(\hat{\beta} - \beta_0) \rightarrow N\left(0, \sum_0^{-1} \sum_w \{\sum_0^{-1}\}^T\right),$$

where β_0 is the true value of β , and $\sum_0$ is the probability limits of $\sum_w$, and $\sum_w = E\{\mathbf{W}_i \mathbf{W}_i^T\}$.

The proof can be easily derived from the derivation in the Appendix. To make inferences, we consistently estimate the matrix $\sum_0$ and $\sum_w$ using $\widehat{\sum}_0 = n^{-1} \sum_{i=1}^n \widehat{\mathbf{D}}_i \widehat{\mathbf{V}}_i^{-1} \widehat{\mathbf{D}}_i$ and $\widehat{\sum}_w = n^{-1} \sum_{i=1}^n \widehat{\mathbf{W}}_i \widehat{\mathbf{W}}_i^T$, where $\widehat{\mathbf{D}}_i$, $\widehat{\mathbf{V}}_i$, and $\widehat{\mathbf{W}}_i$ are $\mathbf{D}_i$, $\mathbf{V}_i$, and $\mathbf{W}_i$ with unknown parameters replaced by their estimates, respectively.

We note that the preceding theorem treats the working regression model parameter β as known. The simulation study in Section 4 shows that the asymptotic variance estimator that ignores uncertainty in β overestimates the variance. For more explanation, please see [14]. In applications, we estimate the parameter β through solving (13), and we should take the variability of β into account. At this situation, the derivation of the asymptotic results is too complicated, which is beyond the scope of this paper. In practice, we suggest using a bootstrap method to estimate the standard errors of β . In general, the bootstrap method works well.

4. Simulation studies

This section presents the performance of the proposed method compared with other methods commonly used in practice. Specifically, we compare three methods: Method 1 is the proposed method where β is estimated on the basis of (13); Method 2, denoted as OLS, is the method that β is estimated through the estimating equation (7); and Method 3 is the complete case mixed effects model with random intercepts. For each method, we consider two types of responses: continuous and binary. We use exchangeable working correlation matrix for Methods 1 and 2.

Throughout the simulation study, we assume a monotone missing pattern, and the true missing data model is assumed to be

$$\text{logit } \lambda_{ij} = \alpha_0 + \alpha_1 S_{i,j-1} + \alpha_2 S_{ij}, \quad (14)$$

where we set $\alpha_0 = 0$, $\alpha_1 = \log(1.5)$, and we vary α_2 to obtain different missing proportions.

We first consider the continuous outcome, where the true model $E(Y_{ij}|X_{ij}) = \beta_0 + \beta_1 X_{ij} = 1 + X_{ij}$. X_{ij} is generated from $\text{Bin}(1, 0.5)$, Y_{ij} is normally distributed with unit variance, and $\text{corr}(Y_{ij}, Y_{ij}) = 1$ for $J = j$. For the surrogate process, we assume $E(S_{ij}|X_{ij}) = \beta_0 + \beta_1 X_{ij} = 1 + X_{ij}$, where S_{ij} is normally distributed with standard deviation 0.5, $\text{corr}(S_{ij}, Y_{ij}) = 0$, $\text{corr}(S_{ij}, S_{ij}) = 0$ for $J = j$. Because all $\mathbf{Y}$ and $\mathbf{S}$ given $\mathbf{X}$ are normally distributed, it follows that conditional on $(\mathbf{S}, \mathbf{X})$, $\mathbf{Y}$ is also normally distributed with $E(\mathbf{Y}|\mathbf{S}, \mathbf{X})$ a linear form of $\mathbf{S}$ and $\mathbf{X}$. We consider two working regression models for $E(\mathbf{Y}|\mathbf{S}, \mathbf{X})$:

$$\varphi(S_{ij}, X_{ij}; \gamma) = \gamma_0 + \gamma_1 S_{ij}^k + \gamma_2 X_{ij}, \quad k=1, 2.$$

The working regression model is correctly specified when $k=1$ and is incorrectly specified when $k=2$. For the missing data model, we also consider two cases: (i) the missing data

model is correctly specified by model (14), and (ii) the missing data model is misspecified as

$$\text{logit } \lambda_{ij} = \alpha_0^* + \alpha_1^* S_{i,j-1}.$$

We record the results in Tables I and II for sample sizes 200 and 400, where SD is standard deviation of the estimate based on 2000 simulations, MDM means the missing data model, WRM means the working regression model, C means correctly specified, and I means incorrectly specified. Not surprisingly, both the proposed and OLS methods provide DR estimate: When either the missing data model or the working regression model is correctly specified, all the two estimators show negligible empirical biases. However, the proposed method exhibits efficiencies not worse or better than the OLS method in all settings. Particularly, in some scenarios, the proposed method gives huge efficiency improvement than the OLS method. For example, when the sample size is 200, the missing proportion is about 50%, $\rho = 0.8$, the missing data model is correctly specified, and the working regression model is misspecified; the variance ratio of the estimates between the proposed method and the OLS method is 0.27 for β_0 and 0.143 for β_1 . In general, the efficiency gain depends on the missing proportions: As the missing proportion increases, the efficiency gain increases. The efficiency gain also depends on the correlation coefficient (ρ) and sample sizes: As the correlation coefficient increases, the efficiency gain increases, and as the sample size decreases, the efficiency gain increases. Furthermore, when both the missing data and working regression models are misspecified, the proposed method also performs better in terms of the standard deviation. The complete case mixed effects model yields biased results.

To evaluate the performance of the covariance formula derived in Theorem 1, we report the empirical coverage probability for the nominal level 95% in Table III. It can be seen that, when either the missing data model or/and the working regression model is correctly specified, almost all the coverage probabilities are 100%. It is a common problem how to estimate the variance accurately for missing data problem, especially in the case ρ_{ij} is too close to 0. This is due to the fact that the asymptotic variance involves terms such that

R_{ij}/π_{ij}^2 . Even though $n^{-1} \sum_i R_{ij}/\pi_{ij}^2$ converges to $E(R_{ij}/\pi_{ij}^2)$ theoretically, the numeric results have big difference. Qin, Shao, and Zhang [27] also pointed out that it is not very reliable to use the asymptotic variance formula to estimate variance. As an alternative, the bootstrap resampling method was suggested.

We also conducted a simulation study in which both X and S are cluster constant. Table IV lists the results for sample size 200. The proposed method yields consistent and more efficient estimate than the OLS method when either the missing data model or the working regression model is correctly specified.

Second, we consider a model that Y is binary given $\mathbf{X}$, where the true model is

$$\text{logit } \mu_{ij} = \beta_0 + \beta_1 X_{ij}.$$

For the surrogate process S , we assume it is normally distributed with mean $E(S_{ij} | Y_{ij}, X_{ij}) = 1 + Y_{ij} + X_{ij}$ and unit variance. We assume a working regression model

$$\text{logit } \varphi(S_{ij}, X_{ij}) = \gamma_0 + \gamma_1 S_{ij} + \gamma_2 X_{ij}.$$

It can be easily shown that $E(Y_{ij}|S_{ij}, X_{ij}) \neq (S_{ij}, X_{ij})$; that is, the working regression model is incorrect. We list the results in Table V. As expected, the proposed method yields more efficient estimate than the OLS method. However, in this simulation study, the efficiency gain is less compared with the continuous outcome case.

All the previous results suggest that the proposed method provides more efficient estimators than the OLS method in all scenarios when either the missing data model or the working regression model is correctly specified, no matter the outcome is continuous or binary. When both the missing data and working regression models are misspecified, the proposed and the OLS methods yield more similar biases, but the proposed method exhibits much smaller standard deviations compared with the OLS method.

5. Application to a dementia study

We apply the proposed method to the National Alzheimer's Coordinating Center Uniform Dataset. One of the goals of the study is to investigate the factors that influence the occurrence of Alzheimer's disease (AD). The outcome we are using is the clinical diagnosis for subjects as having AD (Yes/No), because the gold standard for AD is through neuropathological examination, which is not available for the majority of subjects. Literature studies show that potential risk factors that may influence the occurrence of AD include sex, family history of dementia (fhdm, yes/no), education (years), and age. The surrogate process outcome, S , is the MMSE score, which is highly correlated with the AD. There are 952 subjects from 29 Alzheimer's Disease Centers included at the entry of this study in 2004. Follow-up visits for subjects are scheduled at approximately 1-year intervals, with up to four clinical follow-up visits until 2008. However, patients may miss a clinic visit or refuse to undergo a clinical examination during the clinic visit, leading to incomplete outcomes, but the annual questionnaire-based MMSE score is fully observed. The missing proportion for the outcome is about 31.2%.

Consider the following regression model for the outcome process

$$\text{logit } \mu_{ij} = \mathbf{X}_{ij}^T \beta, \quad (15)$$

where $\mathbf{X}_{ij}$ is the covariate vector at time point j , which includes sex, fhdm, education, and age. For the correlation, we are using an unstructured working correlation matrix.

We consider the following two working regression models for $\text{logit } E(Y_{ij}|S_{ij}, \mathbf{X}_{ij})$, which are

$$\gamma_0 + \gamma_1 \cdot S_{ij} + \gamma_2^T \mathbf{X}_{ij} \quad \text{and} \quad \gamma_0 + \gamma_1 \cdot \sqrt{S_{ij} + \gamma_2^T \mathbf{X}_{ij}},$$

respectively. We denote them as 'Working regression model 1' and 'Working regression model 2'.

For the missing indicators, we build the following regression model:

$$\text{logit } \lambda_{ij} = \mathbf{u}_{ij}^T \alpha, \quad (16)$$

where $\mathbf{u}_{ij}$ include sex, fhdm, MMSE, education, and age.

In lines with the simulation study, we consider three methods. The first is the proposed method where β is estimated from solving (13), the second is the OLS method where β is estimated from solving (7), and the third is the complete case mixed effects model with random intercepts. The results are reported in Table VI, where SE is the standard deviation based on 500 bootstrap resamples (for the proposed and OLS methods). It can be seen that the proposed method yields similar results as the OLS method, but they are different from the mixed effects model results. The proposed method yields a little more efficient estimates than the OLS method, but the efficiency gain. Both the proposed and OLS methods reveal the same conclusion: Holding constant the effects of the other measures, women are less likely to have AD compared with men; people are easy to have AD if they have family history of dementia; education has a positive effect on preventing the occurrence of AD, and age has a negative effect on the occurrence of AD.

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 in Table VII. It is seen that sex, MMSE, and education are significant.

6. Summary and discussion

This paper considers an estimating equation method to estimate the nuisance parameter γ in the working regression model when the missing data model is correctly specified. This method yields the optimal estimate in the sense of minimizing the trace of the asymptotic covariance matrix of the parameters of interest. Simulation studies demonstrate that the proposed method provides more efficient estimate than the method that γ is estimated on the basis of the general OLS method when the missing data model or/and the working regression model is correctly specified. In many situations, the proposed method can yield substantial efficiency improvement than the OLS method. When both the missing data model and the working regression model are misspecified, the proposed method also provide better results in the sense of standard errors. Overall, the proposed method has superior finite sample performance compared with the OLS method across all the range of the settings. Furthermore, the proposed method has a simple estimating equation form, which is easy to be implemented in practice.

It can be seen that the resultant estimating equations for the parameter γ are different when the missing data parameter is known and unknown. There is an additional term added to the estimating equation when the missing data parameter is estimated, which is to account for the variation induced from the missing data model.

In this paper, we only consider the minimum trace of the asymptotic covariance matrix of the parameters of interest. We can adjust this criteria on the basis of different concerns. For example, in a regression model that the interest lies in a particular parameter, we can naturally adopt the criteria to minimizing the asymptotic variance of this parameter estimator in the same spirit of the proposed method; when there are more than one parameters of interest, we may adjust the criteria by minimizing the sum of these asymptotic variances of these parameters. In general, one may develop a weighted minimum trace criteria if different parameters are not equally treated. One can develop other relevant criteria on the basis of particular concerns; for example, one may also use the determinant of the covariance matrix instead of the trace.

We focussed here primarily on estimation and inference regarding the outcome that is subject to missing. The surrogate process and/or covariates subject to missing are also very

common in practice. A future research is to extend this method to the missing outcome, missing surrogate process, or missing covariates problem. The idea is that we adopt a modified estimating equation as proposed in Chen, Yi, and Cook [28]. This is still under investigation.

Acknowledgments

The authors thank the two referees and the editor for their constructive comments and suggestions.

Dr. Chen were supported in part by the National Institute on Aging grant U01AG016976. This paper presents the findings and conclusions of the authors.

References

1. Little, RJA.; Rubin, DB. Statistical Analysis with Missing Data. 2. John Wiley and Sons Inc; New York: 2002.
2. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73:13–22.
3. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*. 1995; 90:106–121.
4. De Gruttola V, Wulfsohn M, Fischl MA, Tsiatis AA. Modeling the relationship between survival and CD-4 lymphocytes in patients with AIDS and AIDS-related complex. *Journal of Acquired Immune Deficiency Syndromes*. 1993; 125:605–613.
5. Wittes J, Lakatos E, Probstfield J. Surrogate endpoints in clinical trials: cardiovascular disease. *Statistics in Medicine*. 1989; 8:415–425. [PubMed: 2727465]
6. Begg C, Leung D. On the use of surrogate endpoints in randomized trials. *Journal of the Royal Statistical Society, Series A*. 2000; 163:15–27.
7. Leung D. Statistical methods in surrogate endpoints. *Journal of the Royal Statistical Society, Series A*. 2001; 164:485–503.
8. Baker SG, Izmirlian G, Kipnis V. Resolving paradoxes involving surrogate endpoints. *Journal of the Royal Statistical Society, Series A*. 2005; 168:753–762.
9. Burzykowski, T.; Molenberghs, G.; Buyse, M. *The Evaluation of Surrogate Endpoints*. Springer; New York: 2003.
10. Baker SG. Surrogate endpoints: wishful thinking or reality? (Editorial). *Journal of the National Cancer Institute*. 2006; 98(8):502–503. [PubMed: 16622112]
11. Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion). *Journal of the American Statistical Association*. 1999; 94:1096–1120.
12. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics*. 2005; 61:962–972. [PubMed: 16401269]
13. Tsiatis, AA. *Semiparametric Theory and Missing Data*. Springer-Verlag; New York: 2006.
14. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressor are not always observed. *Journal of the American Statistical Association*. 1994; 89:846–866.
15. Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in Medicine*. 2004; 23:2937–2960. [PubMed: 15351954]
16. Carpenter J, Kenward M, Vansteelandt S. A comparison of multiple imputation and inverse probability weighting for analyses with missing data. *Journal of the Royal Statistical Society, Series A*. 2006; 169:571–584.
17. Davidian M, Tsiatis AA, Leon S. Semiparametric estimation of treatment effect in a pretest-posttest study without missing data. *Statistical Science*. 2005; 20:261–301. [PubMed: 19081743]
18. Kang JDY, Schafer JL. Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*. 2007; 22:523–539.

19. Van der Laan, MJ.; Robins, JM. Unified Methods for Censored Longitudinal Data and Causality. Springer; New York: 2003.
20. Seaman S, Copas A. Doubly robust generalized estimating equations for longitudinal data. *Statistics in Medicine*. 2009; 28:937–955. [PubMed: 19153970]
21. Chen B, Zhou XH. Doubly robust estimates for binary longitudinal data analysis with missing response and missing covariates. *Biometrics*. 2011; 67:830–842. [PubMed: 21281272]
22. Robins, JM.; Rotnitzky, A. Comment on “Inference for Semiparametric Models: Some Questions and Answer. In: Bickel, PJ.; Kwon, J., editors. *Statistical Sinica*. Vol. 11. 2001. p. 920-936.
23. Cao WH, Tsiatis AA, Davidian M. Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika*. 2009; 96:723–734. [PubMed: 20161511]
24. Duan X, Qin J, Wang Q. Optimal estimation in surrogate outcome regression problems. *Canadian Journal of Statistics*. 2010; 38:1–14.
25. Tsiatis AA, Davidian M, Cao W. Improved doubly robust estimation when data are monotonely coarsened, with application to longitudinal studies with dropout. *Biometrics*. 2011; 67(2):536–545. [PubMed: 20731640]
26. Pepe MS, Anderson GL. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics, Simulation and Computation*. 1994; 23:939–951.
27. Qin J, Shao J, Zhang B. Efficient and doubly robust imputation for covariate-dependent missing responses. *Journal of the American Statistical Association*. 2008; 103:797–810.
28. Chen B, Yi GY, Cook RJ. Weighted generalized estimating functions for longitudinal response and covariate data that are missing at random. *Journal of the American Statistical Association*. 2010; 105:336–353.

Appendix: Regularity conditions and derivation of the optimal estimating equations

Regularity conditions

The regularity conditions required in this paper include standard conditions that are assumed for the estimating function theory, plus the requirement for the missing data processes and covariate process. Here, we only list the key conditions:

1. $(\mathbf{Y}_i, \mathbf{X}_i, \mathbf{S}_i)$, $i = 1, \dots, n$, are independent and identically distributed random variable vectors.
2. The marginal mean model μ is correctly specified. This condition ensures the consistency of the estimators.
3. $P(R_{ij} = 1 | \mathbf{X}_i, \mathbf{S}_i) > 0$ for all $\mathbf{x}$ in its domain, $j = 1, \dots, J$, for some $\mathbf{x}$. This condition ensures that the estimating functions for μ are bounded, which is necessary for a $\sqrt{n}$ -consistent estimator.
4. μ_0 , μ_1 , and μ_2 lie in the interior of compact sets of $\mathbf{x}$, $\mathbf{s}$, and $\mathbf{t}$.
5. The matrices μ_0 and $\mathbf{I}(\mu)$ are nonsingular.

Other routine conditions are similar to those in Robins, Rotnitzky, and Zhao [3] with a proper modification.

Derivative of the optimal estimating equation when the missing probability is known up to a parameter

We first derive the asymptotic property of $\hat{\alpha}$. Subject to previous regularity conditions, and by usual Taylor expansion of (6) with respect to (α, β) around the true value (α_0, β_0) , we have

$$0 = n^{-1/2} \sum_{i=1}^n \mathbf{D}_i \mathbf{V}_i^{-1} \left[\Delta_i(\hat{\alpha}) (Y_i - \mu_i(\mathbf{X}_i; \hat{\beta})) + (\mathbf{I} - \Delta_i(\hat{\alpha})) (\varphi_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma}) - \mu_i(\mathbf{X}_i; \hat{\beta})) \right] \\ = n^{-1/2} \sum_{i=1}^n \mathbf{D}_i \mathbf{V}_i^{-1} [\Delta_i(\alpha_0)(Y_i - \mu_i(\mathbf{X}_i; \beta_0)) + (\mathbf{I} - \Delta_i(\alpha_0))(\varphi_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma}) - \mu_i(\mathbf{X}_i; \beta_0))] - \sum_0 \left\{ n^{1/2} (\hat{\beta} - \beta_0) \right\} + \Gamma n^{1/2} (\hat{\alpha} - \alpha_0) + o_p(1),$$

where $\hat{\alpha} = (\alpha_1, \dots, \alpha_q)^T$, $q = \dim(\alpha)$, with

$$\Gamma_k = E_{(\mathbf{R}, \mathbf{Y}, \mathbf{S}, \mathbf{X})} [\mathbf{D}_i \mathbf{V}_i^{-1} \Delta_{ik} (Y_i - \varphi_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma}))],$$

and $\Delta_{ik} = \Delta_i(\alpha_0) - \delta_{ik}$, $k = 1, \dots, q$.

By standard Taylor expansion arguments for the missing data score function, we have that

$$n^{1/2} (\hat{\alpha} - \alpha_0) = \mathbf{I}^{-1}(\alpha_0) n^{-1/2} \sum_{i=1}^n \mathbf{U}_{2i}(\alpha_0) + o_p(1). \quad (18)$$

Replacing (18) into (17), we obtain

$$0 = n^{-1/2} \sum_{i=1}^n \mathbf{D}_i \mathbf{V}_i^{-1} [\Delta_i(\alpha_0)(Y_i - \varphi_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma})) + (\mathbf{I} - \Delta_i(\alpha_0))(\varphi_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma}) - \mu_i(\mathbf{X}_i; \beta_0))] \\ + \sum_0 \left\{ n^{1/2} (\hat{\beta} - \beta_0) \right\} - n^{-1/2} \sum_{i=1}^n \Gamma \mathbf{I}^{-1}(\alpha_0) \mathbf{U}_{2i}(\alpha_0) + o_p(1)$$

Provided that $\sum_0^{-1}$ is nonsingular, we have

$$n^{1/2} (\hat{\beta} - \beta_0) = n^{-1/2} \sum_{i=1}^n \sum_0^{-1} \mathbf{W}_i + o_p(1),$$

where

$$\mathbf{W}_i = \mathbf{D}_i \mathbf{V}_i^{-1} [\Delta_i(\alpha_0)(Y_i - \mu_i(\mathbf{X}_i; \beta_0)) + (\mathbf{I} - \Delta_i(\alpha_0))(\varphi_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma}) - \mu_i(\mathbf{X}_i; \beta_0))] + \Gamma \mathbf{I}^{-1}(\alpha_0) \mathbf{U}_{2i}(\alpha_0).$$

Now, we derive the optimal estimating equation for β . In the same spirit of derivation when the missing data parameter is known, we consider to take the variance of $\mathbf{W}_i$ with respect to $\mathbf{R}$ given $(\mathbf{Y}, \mathbf{S}, \mathbf{X})$. It is noted that

$$\begin{aligned}
\mathbf{W}_i &= \mathbf{D}_i \mathbf{V}_i^{-1} \boldsymbol{\Delta}_i (\alpha_0) (\mathbf{Y}_i - \boldsymbol{\varphi}_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma})) + \mathbf{D}_i \mathbf{V}_i^{-1} (\boldsymbol{\varphi}_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma}) - \mu_i(\mathbf{X}_i; \beta_0)) + \boldsymbol{\Gamma} \boldsymbol{\Gamma}^{-1} (\alpha_0) \mathbf{v}_i (\mathbf{R}_i - \lambda_i) \\
&= \mathbf{D}_i \mathbf{V}_i^{-1} \boldsymbol{\Psi}_i \mathbf{d}_i + \boldsymbol{\Gamma} \boldsymbol{\Gamma}^{-1} (\alpha_0) \mathbf{v}_i \boldsymbol{\Delta}_{2i} \mathbf{d}_i + \mathbf{D}_i \mathbf{V}_i^{-1} (\boldsymbol{\varphi}_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma}) - \mu_i(\mathbf{X}_i; \beta_0)) - \boldsymbol{\Gamma} \boldsymbol{\Gamma}^{-1} (\alpha_0) \mathbf{v}_i \lambda_i \\
&= \mathbf{H}_i \mathbf{d}_i + \mathbf{D}_i \mathbf{V}_i^{-1} (\boldsymbol{\varphi}_i(\mathbf{S}_i, \mathbf{X}_i; \hat{\gamma}) - \mu_i(\mathbf{X}_i; \beta_0)) - \boldsymbol{\Gamma} \boldsymbol{\Gamma}^{-1} (\alpha_0) \mathbf{v}_i \lambda_i,
\end{aligned}$$

where $\boldsymbol{\Delta}_i = \text{diag}(\lambda_{ij}, j = 1, \dots, J)$, and $\mathbf{H}_i = \mathbf{D}_i \mathbf{V}_i^{-1} \boldsymbol{\Psi}_i + \boldsymbol{\Gamma} \boldsymbol{\Gamma}^{-1} (\alpha_0) \mathbf{v}_i \boldsymbol{\Delta}_{2i}$.

In the following derivation, for simplicity, we ignore the subscript i if there is no confusion. Taking the variance of $\mathbf{W}$ with respect to $\mathbf{R}$ given $(\mathbf{Y}, \mathbf{S}, \mathbf{X})$, we can show that the part of the covariance of $\mathbf{W}$, which involves the nuisance parameter α_0 , is $E_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})}[\mathbf{H}_1 \mathbf{H}_1^T]$. Now, we take the first derivative of the trace of the asymptotic covariance matrix of $n^{1/2}(\hat{\alpha}_0 - \alpha_0)$, with respect to $\alpha_k, k = 1, \dots, q$, which is

$$\begin{aligned}
\partial \text{tr} \left[\sum_0^{-1} \text{cov}_{(\mathbf{R}, \mathbf{Y}, \mathbf{S}, \mathbf{X})}(\mathbf{W}) \sum_0^{-1T} \right] / \partial \gamma_k &= \partial \text{tr} E_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} \left[\sum_0^{-1} \mathbf{H} \sum_1 \mathbf{H}^T \sum_0^{-1T} \right] / \partial \gamma_k = \text{tr} E \left[\partial \left(\sum_0^{-1} \mathbf{H} \sum_1 \mathbf{H}^T \sum_0^{-1T} \right) / \partial \gamma_k \right] \\
&= 2 \text{tr} E_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} \left[\sum_0^{-1} \partial \mathbf{H} / \partial \gamma_k \sum_1 \mathbf{H}^T \sum_0^{-1T} \right] \\
&= 2 \text{tr} E_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} \left[\partial \left(\sum_0^{-1} \mathbf{D} \mathbf{V}^{-1} \boldsymbol{\Psi} \right) / \partial \gamma_k \sum_1 \mathbf{H}^T \sum_0^{-1T} \right] \\
&= 2 \text{tr} E_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} \left[\sum_0^{-1} \mathbf{D} \mathbf{V}^{-1} \boldsymbol{\Psi}_k \sum_1 \mathbf{H}^T \sum_0^{-1T} \right] \\
&= 2 \text{tr} E_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} \left[\sum_0^{-1} \mathbf{D} \mathbf{V}^{-1} \boldsymbol{\Psi}_k \sum_1 \left(\boldsymbol{\Psi}^T \mathbf{V}^{-1} \mathbf{D}^T + \boldsymbol{\Delta}_2 \mathbf{v}^T \boldsymbol{\Gamma}^{-1} (\alpha_0) \boldsymbol{\Gamma}^T \right) \sum_0^{-1T} \right] \\
&= 2 \text{tr} E_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} \left[\mathbf{V}^{-1} \mathbf{D}^T \sum_0^{-1T} \sum_0^{-1} \mathbf{D} \mathbf{V}^{-1} \boldsymbol{\Psi}_k \sum_1 \boldsymbol{\Psi} \right] + 2 \text{tr} E_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} \left[\sum_0^{-1} \mathbf{D} \mathbf{V}^{-1} \boldsymbol{\Psi}_k \sum_1 \left(\boldsymbol{\Delta}_2 \mathbf{v}^T \boldsymbol{\Gamma}^{-1} (\alpha_0) \boldsymbol{\Gamma}^T \right) \sum_0^{-1T} \right] \\
&= 2 E_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} [\mathbf{A}_k (\mathbf{Y} - \boldsymbol{\varphi}(\mathbf{S}, \mathbf{X}; \gamma))] + 2 \mathbf{B}_k,
\end{aligned}$$

where $\mathbf{B}_k = \text{tr} E_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} \left[\sum_0^{-1} \mathbf{D} \mathbf{V}^{-1} \boldsymbol{\Psi}_k \sum_1 \left(\boldsymbol{\Delta}_2 \mathbf{v}^T \boldsymbol{\Gamma}^{-1} (\alpha_0) \boldsymbol{\Gamma}^T \right) \sum_0^{-1T} \right]$. Let $\mathbf{B} = (\mathbf{B}_1, \dots, \mathbf{B}_q)^T$; then, we have

$$\partial \text{tr} \left[\sum_0^{-1} \text{cov}_{(\mathbf{R}, \mathbf{Y}, \mathbf{S}, \mathbf{X})}(\mathbf{W}) \sum_0^{-1T} \right] / \partial \gamma = 2 E_{(\mathbf{Y}, \mathbf{S}, \mathbf{X})} [\mathbf{A} (\mathbf{Y} - \boldsymbol{\varphi}(\mathbf{S}, \mathbf{X}; \gamma)) + 2 \mathbf{B}].$$

With complete data, the optimal estimating equation for α_0 is

$$\mathbf{0} = \sum_{i=1}^n [\mathbf{A}_i (\mathbf{Y}_i - \boldsymbol{\varphi}_i(\mathbf{S}_i, \mathbf{X}_i; \gamma)) + \mathbf{B}_i],$$

where $\mathbf{A}_i$ and $\mathbf{B}_i$ are realizations of $\mathbf{A}$ and $\mathbf{B}$ for subject i .

Table 1

Empirical bias and standard deviation of $\hat{\beta}$: Both response and surrogate process are continuous with sample size $n = 200$.

			Proposed			OLS			Mixed effects model					
			Bias	SD		Bias	SD		Bias	SD				
MDM	WRM		0	1	0	1	0	1	0	1	0	1	0	1
$z = \log(0.5)$ (50% missing)														
0.8	C	C	-0.063	0.016	0.178	0.161	-0.058	0.022	0.213	0.239	-0.705	-0.021	0.192	0.125
	C	I	-0.065	0.002	0.197	0.185	-0.007	-0.008	0.377	0.489				
	I	C	-0.060	0.016	0.155	0.157	-0.089	0.003	0.225	0.333				
	I	I	-0.172	0.029	0.178	0.171	-0.158	0.049	0.243	0.252				
0.4	C	C	-0.035	0.007	0.108	0.097	-0.023	0.004	0.134	0.140	-0.470	-0.016	0.224	0.139
	C	I	-0.024	-0.007	0.148	0.145	-0.008	-0.001	0.224	0.235				
	I	C	-0.028	0.009	0.121	0.092	-0.028	0.014	0.164	0.162				
	I	I	-0.152	-0.009	0.150	0.134	-0.183	0.012	0.376	0.541				
0.0	C	C	-0.005	-0.006	0.160	0.147	-0.007	-0.009	0.207	0.218	-0.239	-0.032	0.174	0.186
	C	I	0.004	-0.002	0.159	0.158	-0.008	-0.009	0.236	0.229				
	I	C	0.001	0.003	0.141	0.132	0.007	-0.004	0.207	0.196				
	I	I	0.006	-0.005	0.144	0.135	0.008	-0.003	0.244	0.175				
$z = \log(1.5)$ (25% missing)														
0.8	C	C	0.002	0.004	0.078	0.065	0.003	0.004	0.080	0.070	0.268	-0.031	0.111	0.054
	C	I	0.003	0.006	0.081	0.069	0.004	0.002	0.082	0.078				
	I	C	0.025	0.009	0.076	0.063	0.026	0.007	0.076	0.064				
	I	I	0.060	0.003	0.078	0.067	0.049	0.002	0.078	0.067				
0.4	C	C	0.003	0.002	0.056	0.046	0.000	0.003	0.057	0.046	0.164	-0.033	0.076	0.050
	C	I	0.002	0.002	0.065	0.055	0.003	-0.004	0.065	0.062				
	I	C	0.048	0.007	0.059	0.043	0.049	0.008	0.059	0.043				
	I	I	0.054	0.005	0.065	0.054	0.054	0.008	0.065	0.057				
0.0	C	C	0.000	-0.002	0.057	0.056	0.000	0.001	0.057	0.056	0.079	-0.024	0.063	0.063
	C	I	-0.001	0.002	0.056	0.056	-0.001	0.002	0.056	0.056				
	I	C	-0.002	0.000	0.054	0.053	0.001	-0.002	0.053	0.055				
	I	I	0.000	0.003	0.053	0.054	-0.001	-0.001	0.054	0.055				

MDM, missing data model; WRM, working regression model; C, correctly specified; I, incorrectly specified; SD, standard deviation; OLS, ordinary least square.

Table II

Empirical bias and standard deviation of $\hat{\beta}$: Both response and surrogate process are continuous with sample size $n = 400$.

	MDM	WRM	Proposed				OLS				Mixed effects model			
			Bias		SD		Bias		SD		Bias		SD	
			0	1	0	1	0	1	0	1	0	1	0	1
$z = \log(0.5)$ (50% missing)														
0.8	C	C	-0.052	0.016	0.147	0.134	-0.040	0.023	0.184	0.224	-0.699	-0.034	0.150	0.085
	C	I	-0.044	0.018	0.160	0.149	-0.002	0.009	0.328	0.297				
	I	C	-0.055	0.014	0.121	0.118	-0.083	0.015	0.160	0.200				
	I	I	-0.181	0.045	0.130	0.132	-0.161	0.059	0.172	0.242				
0.4	C	C	-0.027	0.005	0.091	0.078	-0.015	0.002	0.120	0.131	-0.480	-0.024	0.137	0.087
	C	I	-0.028	-0.006	0.122	0.111	0.004	-0.010	0.405	0.660				
	I	C	-0.023	0.009	0.093	0.075	-0.026	0.027	0.152	0.167				
	I	I	-0.163	0.002	0.116	0.114	-0.189	0.020	0.176	0.220				
0.0	C	C	0.000	0.004	0.128	0.118	-0.010	0.002	0.182	0.200	-0.252	-0.036	0.120	0.108
	C	I	-0.006	0.005	0.116	0.110	-0.004	-0.004	0.203	0.189				
	I	C	-0.005	0.001	0.111	0.098	0.006	0.003	0.150	0.123				
	I	I	0.002	0.003	0.109	0.105	-0.001	0.004	0.139	0.153				
$z = \log(1.5)$ (25% missing)														
0.8	C	C	0.002	0.002	0.058	0.046	-0.001	-0.001	0.059	0.047	0.253	-0.033	0.069	0.037
	C	I	0.003	0.003	0.058	0.051	0.001	0.004	0.058	0.053				
	I	C	0.023	0.005	0.053	0.046	0.028	0.006	0.053	0.048				
	I	I	0.049	0.005	0.057	0.048	0.047	-0.003	0.058	0.048				
0.4	C	C	0.001	0.000	0.040	0.033	0.001	0.000	0.040	0.033	0.166	-0.029	0.056	0.043
	C	I	0.003	0.001	0.044	0.038	0.001	0.001	0.045	0.042				
	I	C	0.050	0.008	0.044	0.031	0.048	0.008	0.044	0.032				
	I	I	0.053	0.007	0.047	0.037	0.054	0.007	0.047	0.038				
0.0	C	C	0.000	-0.001	0.038	0.040	0.000	0.002	0.038	0.040	0.082	-0.028	0.039	0.045
	C	I	-0.001	0.003	0.040	0.038	0.000	0.000	0.040	0.039				
	I	C	0.000	-0.004	0.037	0.038	0.002	-0.001	0.037	0.038				
	I	I	-0.001	-0.001	0.039	0.037	0.002	0.001	0.039	0.037				

MDM, missing data model; WRM, working regression model; C, correctly specified; I, incorrectly specified; SD, standard deviation; OLS, ordinary least square.

Table III
Empirical 95% coverage probability of : Both response and surrogate process are continuous.

<i>n</i>	<i>e</i>	(MDM, WRM)		CC		CI		IC		II	
		0	1	0	1	0	1	0	1	0	1
200	0.8	0.5	1.000	1.000	1.000	0.967	1.000	1.000	1.000	0.364	0.232
		1.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.800	0.560
	0.4	0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.378	0.408
		1.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.800	0.770
	0.0	0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.571	0.469
		1.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.860	0.810
400	0.8	0.5	1.000	1.000	1.000	0.989	1.000	1.000	1.000	0.310	0.200
		1.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.810	0.370
	0.4	0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.240	0.330
		1.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.720	0.770
	0.0	0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.450	0.500
		1.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.830	0.850

MDM, missing data model; WRM, working regression model; C, correctly specified; I, incorrectly specified.

Empirical bias and standard deviation of $\hat{\beta}$: Both X and S are cluster constant, and the response and surrogate process are continuous with sample size $n = 200$.

Table IV

		Proposed				OLS				Mixed effects model			
		Bias		SD		Bias		SD		Bias		SD	
MDM	WRM	0	1	0	1	0	1	0	1	0	1	0	1
$z = \log(0.5)$ (50% missing)													
0.8	C	-0.002	-0.006	0.174	0.083	-0.009	-0.001	0.210	0.285	-1.630	-1.000	0.185	0.043
	C	-0.008	-0.008	0.211	0.081	0.035	-0.011	0.265	0.249				
	I	-0.006	-0.009	0.142	0.064	-0.004	-0.013	0.208	0.402				
	I	-0.654	-0.246	0.193	0.118	-0.599	-0.276	0.205	0.281				
0.4	C	-0.007	-0.007	0.174	0.123	-0.006	-0.001	0.216	0.182	-1.546	-0.972	0.191	0.067
	C	-0.003	-0.006	0.197	0.149	-0.003	-0.004	0.310	0.211				
	I	-0.008	-0.001	0.164	0.106	-0.002	-0.007	0.263	1.049				
	I	-0.491	-0.815	0.167	0.137	-0.562	-0.842	0.182	0.173				
0.0	C	-0.002	-0.001	0.192	0.164	-0.002	-0.003	0.198	0.191	-1.472	-0.956	0.177	0.089
	C	-0.002	-0.008	0.181	0.170	-0.007	-0.006	0.417	0.627				
	I	-0.005	-0.004	0.137	0.143	-0.003	-0.006	0.240	0.480				
	I	-0.002	-0.006	0.130	0.138	-0.005	-0.008	0.155	0.313				
$z = \log(1.5)$ (25% missing)													
0.8	C	-0.002	-0.005	0.140	0.092	-0.004	-0.004	0.170	0.160	0.941	-0.970	0.174	0.035
	C	-0.008	-0.003	0.142	0.084	-0.004	-0.004	0.159	0.244				
	I	0.009	-0.004	0.118	0.073	0.004	-0.001	0.185	0.238				
	I	0.290	-0.963	0.160	0.169	0.275	-0.957	0.172	0.178				
0.4	C	-0.009	-0.005	0.148	0.117	-0.002	-0.001	0.153	0.148	0.821	-0.889	0.138	0.072
	C	-0.009	-0.005	0.160	0.126	-0.006	-0.009	0.238	0.164				
	I	0.005	-0.003	0.125	0.110	0.006	-0.005	0.138	0.155				
	I	0.147	-0.603	0.128	0.138	0.112	-0.740	0.134	0.145				
0.0	C	-0.006	-0.001	0.137	0.138	-0.001	-0.009	0.147	0.179	0.703	-0.781	0.149	0.113
	C	-0.007	-0.006	0.133	0.168	-0.003	-0.009	0.145	0.170				
	I	-0.001	-0.008	0.125	0.118	-0.003	-0.008	0.159	0.174				

		Proposed			OLS			Mixed effects model		
		Bias	SD		Bias	SD		Bias	SD	
MDM	WRM	0	1	0	1	0	1	0	1	0
I	I	-0.008	-0.002	0.120	0.125	-0.061	-0.009	0.123	0.127	

MDM: missing data model; WRM: working regression model; C: correctly specified; I: incorrectly specified; SD, standard deviation; OLS, ordinary least square.

Table V

Empirical bias and standard deviation of : The outcome is binary, and the surrogate outcome is continuous.

n	MDM	WRM	Proposed				OLS				Mixed effects model														
			Bias		SD		Bias		SD																
			0	1	0	1	0	1	0	1	0	1	0	1	0	1									
200	0.8	C	I	-0.056	0.087	$\pi_2 = \log(0.5)$ (50% missing)																			
						I	-0.192	0.099	0.172	0.174	-0.194	0.095	0.175	0.185	0.090	-1.596	1.704	0.912							
		0.4	C	I	-0.004														-0.002	0.162	0.201	0.000	0.005	0.171	0.221
						I	-0.077	0.018	0.158	0.177	-0.071	0.003	0.169	0.204											
															0.0	C	I	0.004							
	I	-0.015	0.029	0.141	0.189	0.000	-0.002	0.149	0.210																
	400	0.8	C	I	-0.061	0.062	$\pi_2 = \log(1.5)$ (25% missing)																		
							I	0.003	0.113	0.161	0.150	0.006	0.111	0.159	0.155	3.561	-1.536	1.442	0.575						
			0.4	C	I	-0.001														-0.004	0.151	0.180	0.005	-0.007	0.153
							I	0.018	0.016	0.147	0.159	0.040	-0.004	0.148	0.179										
0.0																C	I	-0.002	-0.003						
		I	-0.019	0.031	0.130	0.174	0.001	0.000	0.134	0.194															
600		0.8	C	I	-0.056	0.041	$\pi_2 = \log(0.5)$ (50% missing)																		
							I	-0.185	0.087	0.122	0.127	-0.189	0.103	0.120	0.130	-0.218	-1.429	1.280	0.461						
			0.4	C	I	0.005														0.000	0.118	0.150	0.010	-0.014	0.120
							I	-0.069	-0.003	0.114	0.131	-0.074	0.002	0.117	0.144										
	0.0															C	I	0.004	0.000						
		I	-0.004	0.007	0.103	0.142	0.004	-0.004	0.108	0.151															
	800	0.8	C	I	-0.051	0.045	$\pi_2 = \log(1.5)$ (25% missing)																		
							I	-0.001	0.112	0.109	0.108	0.004	0.113	0.117	0.111	3.405	-1.604	1.120	0.433						
			0.4	C	I	0.003														-0.002	0.111	0.132	0.004	-0.005	0.110
							I	0.030	0.001	0.107	0.123	0.030	-0.002	0.110	0.127										

		Proposed				OLS				Mixed effects model			
		Bias		SD		Bias		SD		Bias		SD	
<i>n</i>		MDM	WRM	0	1	0	1	0	1	0	1	0	1
0.0	C	I		0.000	-0.002	0.095	0.135	-0.001	0.000	0.096	0.135	0.157	-0.011
	I	I		-0.001	0.008	0.094	0.133	0.002	0.003	0.097	0.144	0.126	0.188

MDM, missing data model; WRM, working regression model; C, correctly specified; I, incorrectly specified; SD, standard deviation; OLS, ordinary least square.

Point estimate for the National Alzheimer's Coordinating Center Uniform Dataset: outcome regression model (the correlation matrix is unstructured).

Table VI

Parameter	Proposed			OLS			Mixed effects model		
	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
Working regression model 1:									
(Intercept)	-1.797	0.131	<0.001	-1.839	0.139	<0.001	-2.107	2.992	0.481
SEX(F)	-0.216	0.026	<0.001	-0.159	0.022	<0.001	-0.778	0.529	0.141
fhdem	0.182	0.072	0.011	0.186	0.075	0.013	0.690	0.363	0.057
Education	-0.007	0.002	<0.001	-0.007	0.002	<0.001	-0.007	0.024	0.764
Age	0.031	0.002	<0.001	0.031	0.002	<0.001	0.074	0.038	0.051
Working regression model 2:									
(Intercept)	-1.779	0.122	<0.001	-1.820	0.153	<0.001			
SEX	-0.213	0.024	<0.001	-0.153	0.023	<0.001			
fhdem	0.184	0.070	0.004	0.189	0.071	0.004			
Education	-0.007	0.002	0.007	-0.007	0.002	0.001			
Age	0.030	0.002	<0.001	0.031	0.002	<0.001			

fhdem, family history of dementia; SE, standard error; OLS, ordinary least square.

Table VII

Point estimate for the National Alzheimer's Coordinating Center Uniform Dataset: missing data model.

Parameter	Estimate	SE	<i>p</i>
(Intercept)	24.318	2.059	<0.001
SEX	−0.760	0.169	<0.001
fhdem	−0.244	0.177	0.169
Education	−0.025	0.009	0.003
Age	−0.023	0.012	0.055
MMSE	−0.750	0.057	<0.001

fhdem, family history of dementia; standard error; MMSE, Mini-Mental State Exam.