Published in final edited form as:

Biometrics. 2014 June ; 70(2): 289–298. doi:10.1111/biom.12150.

Test the reliability of doubly robust estimation with missing response data

Baojiang Chen¹ and Jing Qin²

¹Department of Biostatistics, College of Public Health, University of Nebraska Medical Center, Omaha, Nebraska, US 68198, baojiang.chen@unmc.edu

²National Institute of Allergy and Infectious Diseases, National Institute of Health, Bethesda, MD 20892, US, jingqin@niaid.nih.gov

Summary

In statistical inference one has to make sure that the underlying regression model is correctly specified otherwise the resulting estimation may be biased. Model checking is an important method to detect any departure of the regression model from the true one. Missing data is a ubiquitous problem in social and medical studies. If the underlying regression model is correctly specified, recent researches show great popularity of the doubly robust estimates method for handling missing data because of its robustness to the misspecification of either the missing data model or the conditional mean model, i.e. the model for the conditional expectation of true regression model conditioning on the observed quantities. However, little work has been devoted to the goodness of fit test for doubly robust estimates method. In this paper, we propose a testing method to assess the reliability of the estimator derived from the doubly robust estimating equation with possibly missing response and always observed auxiliary variables. Numerical studies demonstrate that the proposed test can control type I errors well. Furthermore the proposed method can detect departures from model assumptions in the marginal mean model of interest powerfully. A real dementia data set is used to illustrate the method for the diagnosis of model misspecification in the problem of missing response with an always observed auxiliary variable for cross-sectional data.

Keywords

Auxiliary; doubly robust; estimating equation; goodness of fit; missing data

1. Introduction

Missing data are common in many settings, and in general, analyses based only on individuals with complete data can lead to invalid inferences when data are missing at random (MAR) or missing not at random (MNAR) (Little and Rubin 2002). Horvitz and Thompson (1952), and Robins, Rotnitzky, and Zhao (1994) developed, respectively, a class

Supplementary Materials

Web Appendices, Tables, and Figures referenced in Sections 4, 5 and 6 are available with this paper at the Biometrics website on Wiley Online Library.

of inverse probability weighted estimating equations (IPWEE) and augmented IPWEE for cross-sectional data 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.

Auxiliary variables are commonly available in many studies (Fleming and DeMets 1996). In this paper, we study the problem where the outcome variable of interest is missing in a subset of the sampled data, but, in addition to the collection of covariate and outcome variables, an auxiliary variable for the outcome is collected for each individual. A typical example for an auxiliary variable is the use of CD4 cell counts in the study of acquired immune deficiency syndrome (De Gruttola et al. 1993). More discussions on auxiliary variables can be found, among others, in Wittes et al. (1989), Begg and Leung (2000), Leung (2001), Baker et al. (2005), Burzykowski et al. (2005) and Baker (2006).

How to use auxiliary variables to improve the robustness and estimation efficiency of the estimate is a challenging problem. In statistical literature, many research papers can be found, in particular the so called doubly robust (DR) methods developed by, Robins, Rotnitzky, and Zhao (1994, 1995), Scharfstein, Rotnitzky and Robins (1999), Bang and Robins (2005), Tsiatis (2006), and among others. In DR estimation, other than the correct specification of the main regression model, two additional models are introduced: (i) the missing data model that describes the missing mechanism and (ii) the conditional mean (or outcome regression) model, i.e., the conditional expectation of true mean conditioning on the observed quantities (see Section 2 for more detailed definition of the conditional mean model in this paper). The resulting DR estimate is robust to misspecification of either the conditional mean model or the missing data model but not both. The comprehensive literature for the doubly robust estimator includes Lunceford and Davidian (2004), Carpenter, Kenward, and Vansteelandt (2006), Davidian, Tsiatis, and Leon (2005), and Kang and Schafer (2007), Cao, Tsiatis and Davidian (2009), Duan, Qin and Wang (2010), Tsiatis, Davidian and Cao (2010), Van der Laan and Robins (2003), Seaman and Copas (2009), Chen and Zhou (2011), among others.

In general, the validity of the doubly robust estimating equations depends on whether the missing data model is correctly specified, or whether the conditional mean model is correctly specified. If both the missing data and the conditional mean models are misspecified, the resultant doubly robust estimating equations deviate in expected value from 0, thus an inconsistent estimator will be yielded. To get a valid inference, it is desirable to assess the unbiasedness of the estimating equations. Diagnostic tools such as residuals have been widely used to assess the appropriateness of a generalized linear model (Su and Wei 1991; Lin, Wei and Ying 2002). However, most methods cannot apply to incomplete data. Moreover, to the best of our knowledge, scant literature exists for developing assessment tools for doubly robust estimating equations with missing data.

The aim of this paper is to test the correctness of the main regression model by using doubly robust estimating equations. The idea is based on testing for the unbiasedness of the doubly robust estimating equations. The augmented residual processes will be constructed in the same spirit of Su and Wei (1991), Lin, Wei and Ying (1993, 2002), and Pan and Lin (2002).

Under the null hypothesis that the doubly robust estimating equations are unbiased, the augmented residual processes will fluctuate about 0. Thus a large absolute value of the augmented residuals leads to the conclusion of model misspecification. This method is appealing in that it provides an integrated test for model misspecification, including the marginal mean of the response model, missing data model, as well as the conditional mean model. Specifically, the proposed method is most useful to detect departures in specification of the marginal mean model, in possible presence of misspecified missing data model or the conditional mean model. Another advantage is that it can be easily implemented in practice. Numerical studies show the proposed test can control type I errors if either the missing data model or the conditional mean model is correctly specified. In contrast, the popular IPWEE method may have inflated type I errors if the missing data model is misspecified.

The paper is organized as follows: we give notations and postulate models in Section 2. In Section 3, we introduce the doubly robust estimate. In Section 4, we propose a lack of fit test for the doubly robust estimating equations, and we evaluate this proposed method by numerical studies in Section 5. In Section 6, we apply this method to an Alzheimer's disease study. Concluding remarks are given in Section 7.

2. Notations and Models

Assume we observe n independent $(y_i, s_i, \mathbf{x}_i)$, $i = 1, \dots, n$, where $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$, $x_{i1} = 1$, denotes a vector of $p \times 1$ fixed covariates for the i th subject, y_i denotes an observation of the outcome random variable Y_i , and s_i denotes an observation of an auxiliary random variable S_i . Let $\mu_i(\beta) = E(Y_i | \mathbf{x}_i; \beta)$ denote the marginal mean of the response Y_i given the covariate $\mathbf{x}_i$, where β is an unknown parameter vector. Typically, we may specify a generalized linear model

$$g(\mu_i) = \mathbf{x}_i^T \beta, \quad (1)$$

to study the association between the response Y_i and the covariate vector $\mathbf{X}_i$, where g is a known monotone differentiable link function.

Similarly, we can build a regression model for the conditional mean of Y_i given s_i and $\mathbf{x}_i$ as $E(Y_i | s_i, \mathbf{x}_i) = \phi_i(s_i, \mathbf{x}_i; \gamma)$, where $\phi_i(\cdot)$ is a given function, and γ is a vector of unknown parameters. We call $\phi_i(\cdot)$ the conditional mean model for simplicity if there is no confusion. Typically, we may specify another generalized linear regression model

$$g_2(\phi_i) = \gamma_0 + \gamma_1 s_i + \gamma_2^T \mathbf{x}_i, \quad (2)$$

where $\gamma = (\gamma_0, \gamma_1, \gamma_2^T)^T$, and g_2 is a link function. If there is no confusion, here and hereafter, we use the notation $\phi_i(\gamma)$ instead of $\phi_i(s_i, \mathbf{x}_i; \gamma)$, for simplicity.

We assume that the response Y_i may be missing, while the covariate $\mathbf{X}_i$ and the auxiliary variable S_i are always observed for every individual. Let R_i be the random missing indicator for Y_i which takes the value 1 if Y_i is observed and 0 otherwise for $i = 1, \dots, n$. Here we assume a missing at random mechanism, where the missing probability for the response is assumed to be

$$P(R_i=1|y_i, s_i, \mathbf{x}_i)=P(R_i=1|s_i, \mathbf{x}_i). \quad (3)$$

Let $\pi(s_i, \mathbf{x}_i; \alpha) = P(R_i = 1|s_i, \mathbf{x}_i)$ be a known function with an unknown parameter vector α . Typically, a logistic regression model is often assumed for π , though other forms may be specified.

In practice, α is unknown, and one may estimate it by maximizing the binomial likelihood

$$\prod_{i=1}^n \{\pi(s_i, \mathbf{x}_i; \alpha)\}^{r_i} \{1 - \pi(s_i, \mathbf{x}_i; \alpha)\}^{1-r_i} \text{ with respect to } \alpha. \text{ Denote the estimate as } \hat{\alpha}.$$

Similarly, the parameter γ is unknown, and one can obtain a valid estimate based on the available data $(y_i, \mathbf{x}_i, s_i, r_i = 1)(i = 1, \dots, n)$, by solving the following estimating equation

$$\sum_{i=1}^n r_i \frac{\partial \phi_i(\gamma)}{\partial \gamma} \{v(\phi_i(\gamma))\}^{-1} (y_i - \phi_i(\gamma)) = 0, \quad (4)$$

where $v(\phi_i(\gamma))$ is a variance function of $\phi_i(\gamma)$. In general, the variance functions can be specified in the sense of generalized linear model. For example, the variance functions $v(\phi)$ are constant σ^2 , $\phi(1 - \phi)$ and ϕ for normal, Bernoulli, and Poisson distributions. Solving (4) will yield a consistent estimate for γ if $\phi(\gamma)$ is correctly specified under the missing at random mechanism (3).

3. Doubly Robust Estimates with Missing Responses

Let $\mathbf{A}(\mathbf{x}; \beta)$ be some $p \times 1$ vector-valued function $\mathbf{x}$ and β . If there is no missing data, we may solve for the parameter β based on the following estimating equation

$\sum_{i=1}^n \mathbf{A}(\mathbf{x}_i; \beta)(y_i - \mu_i(\beta)) = 0$. Under some regularity conditions, this gives a consistent estimate of β if the marginal mean μ_i is correctly specified since the estimating equation has expectation zero. For expression simplicity, here and hereafter we use the term $\mathbf{A}(\mathbf{x})$ instead of $\mathbf{A}(\mathbf{x}; \beta)$ if there is no confusion.

In the presence of missing data, a common method of estimation is the weighted estimating equation (WEE) method proposed by Horvitz and Thompson (1952), which is to solve the following estimating equation

$$\sum_{i=1}^n \mathbf{A}(\mathbf{x}_i) \frac{r_i}{\pi(s_i, \mathbf{x}_i; \hat{\alpha})} (y_i - \mu_i(\beta)) = 0, \quad (5)$$

where $\hat{\alpha}$ is an appropriately chosen and consistent estimator of α . The estimating equation given by (5) is unbiased for 0 at the true β if both the marginal mean μ and the missing data model π are correctly specified. If the marginal mean model μ is correctly specified, but the missing data model is misspecified, the resultant estimator, in general, is inconsistent. To improve the robustness of the estimator, doubly robust estimate method was developed (Robins, Rotnitzky and Zhao 1994; Tsiatis 2006).

By augmenting the weighted estimating equation (5), Robins, Rotnitzky and Zhao (1994) and Tsiatis (2006) used a new estimating equation through

$$\sum_{i=1}^n \mathbf{A}(\mathbf{x}_i) \left[\frac{r_i}{\pi_i(s_i, \mathbf{x}_i; \hat{\alpha})} (y_i - \mu_i(\beta)) + \left(1 - \frac{r_i}{\pi_i(s_i, \mathbf{x}_i; \hat{\alpha})}\right) (\phi_i(\hat{\gamma}) - \mu_i(\beta)) \right] = 0, \quad (6)$$

for estimation of β , where $\hat{\alpha}$ and $\hat{\gamma}$ are some appropriately chosen and consistent estimators of α and γ . The estimating equation is preferred over the weighted estimating equation for its double robustness (Scharfstein, Rotnitzky and Robins 1999; Tsiatis 2006). That is, this estimating equation can produce a consistent estimator of β if either the missing data model π or the conditional mean model ϕ is correctly specified, conditional that the marginal mean μ is correctly specified.

Therefore, it is desirable to assess the validity of the doubly robust estimating equations before making inference. Violation of the unbiasedness of the doubly robust estimating equation may be because of the misspecification of the marginal mean model, or misspecification of both the missing data model and the conditional mean model. This complicates the problem, since it involves testing three models: the marginal mean model μ , the missing data model π and the conditional mean model ϕ . In the following section, we propose an integrated testing method to assess the reliability of the estimator derived from the doubly robust estimating equation, which is easy to implement in practice. Specifically, the proposed method is most useful to detect departures in specification of the marginal mean model, in possible presence of misspecified missing data model or the conditional mean model.

4. Lack of Fit Test with Missing Response

In the absence of missing data, Su and Wei (1991) developed a model checking technique based on the partial sums of the residuals. This method could be used to check the missing data and conditional mean models. The residuals are defined by $e_i = y_i - \mu_i(\beta)$, for $i = 1, \dots, n$. To check the mean models, Su and Wei (1991) studied the process:

$W_n(\mathbf{t}, \hat{\beta}) = n^{-1/2} \sum_{i=1}^n I(\mathbf{x}_i \leq \mathbf{t}) \hat{e}_i$, where $\mathbf{t}$ is a vector of length p , $I(\cdot)$ is an indicator function, and $\hat{e}_i = y_i - \mu_i(\hat{\beta})$. The $W_n(\mathbf{t}; \hat{\beta})$ is a multiparameter stochastic process. Under the null hypothesis that the mean model μ is correctly specified, we would expect that the partial sum process fluctuates about 0. Thus a large value of $G_n = \sup_{\mathbf{t} \in R^p} |W_n(\mathbf{t}; \hat{\beta})|$ leads to the conclusion of model misspecification. To make a goodness of fit test, we need to specify the distribution of $W_n(\mathbf{t}; \hat{\beta})$. Unfortunately, it is hard to specify the explicit distribution form. Su and Wei (1991) proposed a simulation-based method to approximate the null distribution of $W_n(\mathbf{t}; \hat{\beta})$. The idea is as follows.

Suppose $X_1, \dots, X_n$ are independent and identically distributed (iid) variables from $F(X)$, with $\mu = E(X) = 0$ and $\sigma^2 = E(X^2) < \infty$, then based on the central limit theorem, we have

$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \rightarrow N(0, \sigma^2)$. Let $Z_1, \dots, Z_n$ be independent standard normal random variables. Then conditional on the original data

$X_1, \dots, X_n$, $\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i X_i \sim N(0, n^{-1} \sum_{i=1}^n X_i^2) \rightarrow N(0, \sigma^2)$. That is, $n^{-1/2} \sum_{i=1}^n Z_i X_i$

has the same asymptotic distribution as that of $n^{-1/2} \sum_{i=1}^n X_i$. Since the test statistic $W_n(\mathbf{t}; \hat{\beta})$ involves some estimated quantities such as $\hat{\beta}$, we expand it to summation of iid random variables (with mean 0) and an extra higher order remainder (does not impact the asymptotic distribution results) using the standard Taylor expansion. Specifically, applying the standard Taylor expansion to $W_n(\mathbf{t})$ about $\hat{\beta}$ centered at the true value β , we have

$$W_n(\mathbf{t}, \hat{\beta}) = n^{-1/2} \sum_{i=1}^n e_i \{I(\mathbf{x}_i \leq \mathbf{t}) + \eta^T(\mathbf{t}; \beta) \mathbf{\Lambda}(\beta) \mathbf{A}(\mathbf{x}_i)\} + o_p(1), \text{ where}$$

$$\mathbf{\Lambda}^{-1}(\beta) = -n^{-1} \sum_{i=1}^n \partial \mathbf{U}_i / \partial \beta^T, \mathbf{U}_i(\beta) = \mathbf{A}(\mathbf{x}_i)(y_i - \mu_i(\beta)), \eta(\mathbf{t}; \beta) = -n^{-1} \sum_{i=1}^n I(\mathbf{x}_i \leq \mathbf{t}) \partial \mu_i(\beta) / \partial \beta^T$$

Based on the above result, and under the null hypothesis that (1) holds, it follows that

$\tilde{W}_n(\mathbf{t}; \hat{\beta}) = n^{-1/2} \sum_{i=1}^n z_i \hat{e}_i \{I(\mathbf{x}_i \leq \mathbf{t}) + \eta^T(\mathbf{t}; \hat{\beta}) \mathbf{\Lambda}(\hat{\beta}) \mathbf{U}_i(\hat{\beta})\}$ has the same asymptotic distribution as that of $W_n(\mathbf{t}; \beta)$, where $z_1, \dots, z_n$ are independent standard normal variables that independent of $(y_i, \mathbf{x}_i)$, $i = 1, \dots, n$. Furthermore, under the null hypothesis that (1) holds, it follows that $\hat{G}_n = \sup_{\mathbf{t} \in R^p} |W_n(\mathbf{t}; \hat{\beta})|$ has the same asymptotic distribution as that of G_n . To approximate the distribution of G_n , we simulate a number of realizations $W_n(\mathbf{t}; \hat{\beta})$ by repeatedly generating the normal samples $(z_1, \dots, z_n)$ with fixing the data $(y_i, \mathbf{x}_i)$, $i = 1, \dots, n$, at their observed values.

In the presence of missing data, the method of Su and Wei (1991) does not work. In the following, we modify the method of Su and Wei (1991) to the missing response problem. We first consider the assessment of the weighted estimating equation (5), then we further adopt it to assess the doubly robust estimating equation (6).

4.1 Lack of fit test for weighted estimating equations

To check the validity of the weighted estimating equation (5), we consider the stochastic process

$$W_n(\mathbf{t}) = n^{-1/2} \sum_{i=1}^n w_i(\hat{\alpha}) I(\mathbf{x}_i \leq \mathbf{t}) \hat{e}_i, \quad (7)$$

where $w_i(\alpha) = r_i / \pi(s_i, \mathbf{x}_i; \alpha)$. Under the null hypothesis that the weighted estimating equation (5) equals zero in expected value for all values of x_i , one would expect that the partial sum process $W_n(\mathbf{t})$ fluctuates about 0. If the missing data model π is correctly specified and the mean function μ_i is misspecified, the weighted estimating equation (5) will deviate from zero. Thus a large value of $G_n = \sup_{\mathbf{t} \in R^p} |W_n(\mathbf{t})|$ leads to the conclusion of model misspecification. Similarly, misspecification of the missing model π could also lead to large values of G_n .

To make the goodness of fit test, we may also need to find the limiting distribution of $W_n(\mathbf{t})$ given above. In general, it is hard to specify the explicit distribution form. In the same spirit of Su and Wei (1991), however, numerical approximation of the distribution of $W_n(\mathbf{t})$ is possible. Now we give a brief derivation of the numerical approximation of the distribution of $W_n(\mathbf{t})$. For detailed derivation and weak convergence of $W_n(\mathbf{t})$, please see Appendix 1 in the Supplementary Materials.

Under the null hypothesis that both the marginal mean of the response and missing data models are correctly specified, we can obtain a consistent estimator $\hat{\beta}$ by solving (5). Thus, applying the standard Taylor expansion to (7) about $(\hat{\beta}, \hat{\alpha})$ centered at the true value (β, α) , we have

$$W_n(\mathbf{t}) = n^{-1/2} \sum_{i=1}^n w_i(\alpha)(y_i - \mu_i(\beta))I(\mathbf{x}_i \leq \mathbf{t}) + \eta_1^T(\mathbf{t}; \beta, \alpha)n^{1/2}(\hat{\beta} - \beta) + \eta_3^T(\mathbf{t}; \beta, \alpha)n^{1/2}(\hat{\alpha} - \alpha) + o_p(1), \quad (8)$$

where

$$\eta_1(\mathbf{t}, \beta, \alpha) = -n^{-1} \sum_{i=1}^n [\partial \mu_i(\beta) / \partial \beta w_i(\alpha) I(\mathbf{x}_i \leq \mathbf{t})], \quad \eta_3(\mathbf{t}; \beta, \alpha) = n^{-1} \sum_{i=1}^n [\partial w_i(\alpha) / \partial \alpha (y_i - \mu_i(\beta)) I(\mathbf{x}_i \leq \mathbf{t})]$$

Applying the standard Taylor expansion to the weighted estimating equation (5) about $(\hat{\beta}, \hat{\alpha})$ centered at the true value (β, α) and after some algebra, we obtain

$$n^{1/2}(\hat{\beta} - \beta) = \Gamma^{-1} [n^{-1/2} \sum_{i=1}^n \mathbf{A}(\mathbf{x}_i) w_i(\alpha)(y_i - \mu_i(\beta)) + \Gamma_{12} n^{1/2}(\hat{\alpha} - \alpha)] + o_p(1), \quad (9)$$

where $\Gamma = E[\mathbf{A}(\mathbf{x})w(\alpha) \mu'(\beta)^T]$, and $\Gamma_{12} = E[\mathbf{A}(\mathbf{x})(y - \mu(\beta)) w(\alpha) / \alpha^T]$.

Similarly, applying the standard Taylor expansion for the score function of the missing data parameter α about $(\hat{\alpha})$ centered at the true value (α) , we obtain

$$n^{1/2}(\hat{\alpha} - \alpha) = \mathbf{I}_\alpha^{-1} n^{-1/2} \sum_{i=1}^n \mathbf{U}_{3i}(\alpha) + o_p(1), \quad (10)$$

where $\mathbf{I}_\alpha = -E[\mathbf{U}_3(\alpha) / \alpha \mathbf{A}(\mathbf{x})]$, and $\mathbf{U}_{3i}(\alpha)$ is the score function for the missing data parameter α . Plug (9) and (10) into (8), we have

$$W_n(\mathbf{t}) = n^{-1/2} \sum_{i=1}^n \left\{ [I(\mathbf{x}_i \leq \mathbf{t}) + \eta_1^T(\mathbf{t}; \beta, \alpha) \Gamma^{-1} \mathbf{A}(\mathbf{x}_i)] w_i(\alpha) e_i + \{ \eta_1^T(\mathbf{t}; \beta, \alpha) \Gamma^{-1} \Gamma_{12} + \eta_3^T(\mathbf{t}; \beta, \alpha) \} \mathbf{I}_\alpha^{-1} \mathbf{U}_{3i}(\alpha) \right\} + o_p(1). \quad (11)$$

In general, under the null hypothesis that the weighted estimating equation (5) is equal to zero in expected value for all values of x_i , for large n , the distribution of $W_n(\mathbf{t})$ can be approximated by that of $\tilde{W}_n(\mathbf{t}; \hat{\beta}, \hat{\alpha})$, where

$$\tilde{W}_n(\mathbf{t}; \hat{\beta}, \hat{\alpha}) = n^{-1/2} \sum_{i=1}^n \left\{ [I(\mathbf{x}_i \leq \mathbf{t}) + \eta_1^T(\mathbf{t}; \hat{\beta}, \hat{\alpha}) \hat{\Gamma}^{-1} \mathbf{A}(\mathbf{x}_i)] w_i(\hat{\alpha}) \hat{e}_i + \eta_1^T(\mathbf{t}; \hat{\beta}, \hat{\alpha}) \hat{\Gamma}^{-1} \hat{\Gamma}_{12} \hat{\Gamma}_\alpha^{-1} \mathbf{U}_{3i}(\hat{\alpha}) \right\} z_i,$$

$$z_i \sim N(0, 1), \quad \hat{\Gamma} = n^{-1} \sum_{i=1}^n \mathbf{A}(\mathbf{x}_i) w_i(\hat{\alpha}) \partial \mu_i(\hat{\beta}) / \partial \beta^T, \quad \hat{\Gamma}_{12} = n^{-1} \sum_{i=1}^n \mathbf{A}(\mathbf{x}_i) \hat{e}_i \partial w_i(\hat{\alpha}) / \partial \alpha^T,$$

and $\hat{\Gamma}_\alpha = -n^{-1} \sum_{i=1}^n \partial \mathbf{U}_{3i}(\hat{\alpha}) / \partial \alpha^T$.

To approximate the null distribution $W_n(\mathbf{t})$, we obtain a large number of realizations of $\tilde{W}_n(\mathbf{t}; \hat{\beta}, \hat{\alpha})$ by repeatedly generating the normal random sample $(z_1, \dots, z_n)$ while fixing the data $(y_i, \mathbf{x}_i, r_i)$, $i = 1, \dots, n$, at their observed values.

4.2 Lack of fit test for doubly robust estimating equations

We now consider the assessment of the doubly robust estimating equation (6). In the same spirit of augmenting method to construct the doubly robust estimating equation, we consider the statistical process

$$W_n(\mathbf{t}) = n^{-1/2} \sum_{i=1}^n [w_i(\hat{\alpha}) \hat{e}_{1i} + (1 - w_i(\hat{\alpha})) \hat{e}_{2i}] I(\mathbf{x}_i \leq \mathbf{t}), \quad (12)$$

where $\hat{e}_{1i} = y_i - \mu_i(\hat{\beta})$, and $\hat{e}_{2i} = \phi_i(\hat{\gamma}) - \mu_i(\hat{\beta})$. Under the null hypothesis that the doubly robust estimating equation (6) equals zero in expected value for all values of x_i , one would expect that the partial sum process $W_n(\mathbf{t})$ fluctuates about 0. If the marginal mean μ is misspecified, or both the missing data model π and the conditional mean model ϕ are misspecified, the doubly robust estimating equation (6) will deviate from zero. Thus a large value of $G_n = \sup_{\mathbf{t} \in R^p} |W_n(\mathbf{t})|$ leads to violation of the null hypothesis that the doubly robust estimating equation (6) equals zero in expected value for all values of x_i .

Similarly, in the same spirit of derivation in Section 4.1, numerical approximation of the distribution of $W_n(\mathbf{t})$ will be applied. Specifically, under the null hypothesis that the doubly robust estimating equation (6) is equal to zero in expected value for all values of x_i , and for large n , the distribution of $W_n(\mathbf{t})$ can be approximated by that of $\tilde{W}_n(\mathbf{t}; \hat{\beta}, \hat{\alpha}, \hat{\gamma})$, where

$$\begin{aligned} \tilde{W}_n(\mathbf{t}; \hat{\beta}, \hat{\alpha}, \hat{\gamma}) \\ = n^{-1/2} \sum_{i=1}^n \left\{ [I(\mathbf{x}_i \leq \mathbf{t}) + \eta_1^T(\mathbf{t}; \hat{\beta}, \hat{\alpha}, \hat{\gamma}) \hat{\Gamma}^{-1} \mathbf{A}(\mathbf{x}_i)] [w_i(\hat{\alpha}) \hat{e}_{1i} + (1 - w_i(\hat{\alpha})) \hat{e}_{2i}] + \eta_1^T(\mathbf{t}; \hat{\beta}, \hat{\alpha}, \hat{\gamma}) \hat{\Gamma}^{-1} \hat{\Gamma}_{11} \hat{\Gamma}_{\gamma}^{-1} \mathbf{U}_{2i}(\hat{\gamma}) + \eta_1^T(\mathbf{t}; \hat{\beta}, \hat{\alpha}, \hat{\gamma}) \right\} \end{aligned}$$

$$z_i \sim N(0, 1), \quad \hat{e}_{1i} = y_i - \mu_i(\hat{\beta}), \quad \hat{e}_{2i} = \phi_i(\hat{\gamma}) - \mu_i(\hat{\beta}),$$

$$\begin{aligned} \hat{\Gamma} &= n^{-1} \sum_{i=1}^n \mathbf{A}(\mathbf{x}_i) \partial \mu_i / \partial \beta^T |_{\beta=\hat{\beta}}, \quad \hat{\Gamma}_{11} \\ &= n^{-1} \sum_{i=1}^n \mathbf{A}(\mathbf{x}_i) (1 - w_i) \partial \phi_i / \partial \gamma^T |_{(\beta, \alpha, \gamma)=(\hat{\beta}, \hat{\alpha}, \hat{\gamma})}, \quad \hat{\Gamma}_{12} \\ &= n^{-1} \sum_{i=1}^n \mathbf{A}(\mathbf{x}_i) (y_i - \phi_i) \partial w_i / \partial \alpha^T |_{(\beta, \alpha, \gamma)=(\hat{\beta}, \hat{\alpha}, \hat{\gamma})}, \quad \hat{\Gamma}_{\alpha} \\ &= -n^{-1} \sum_{i=1}^n \partial \mathbf{U}_{3i}(\hat{\alpha}) / \partial \alpha^T, \quad \hat{\Gamma}_{\gamma} = -n^{-1} \sum_{i=1}^n \partial \mathbf{U}_{2i}(\hat{\gamma}) / \partial \gamma^T, \text{ and } \eta_1(\mathbf{t}; \hat{\beta}, \hat{\alpha}, \hat{\gamma}) \text{ is} \end{aligned}$$

given in the Appendix 2 in the Supplementary Materials. The proof is given in the Appendix 2 in the Supplementary Materials.

Similarly, to approximate the null distribution $W_n(\mathbf{t})$, we obtain a large number of realizations of $W_n(\mathbf{t}; \hat{\beta}, \hat{\alpha}, \hat{\gamma})$ by repeatedly generating the normal random sample $(z_1, \dots, z_n)$ while fixing the data $(y_i, \mathbf{x}_i, s_i, r_i)$, $i = 1, \dots, n$ at their observed values.

It is noted that the above test statistic G_n serves as an omnibus test, i.e., a global goodness-of-fit test of the marginal mean of response, missing data model, or conditional mean model. If departures are detected, misspecification of models may be from the mean model of interest or both misspecification of the missing data model and of the conditional mean model.

The proposed method can also be adopted to check the link function by considering the partition using the estimated predictor $\mathbf{x}_i^T(\hat{\beta})$ (Lin, Wei and Ying 2002; Hosmer and Lid Hjort 2002); or, to check a specified function form of covariate effect, say x_{ij} , by replacing $I(\mathbf{x}_i = \mathbf{t})$ by $I(x_{ij} = t_j)$ as proposed by Lin, Wei and Ying (2002), which is a special case of (12) with $t_{ik} = \infty$ for all $k \neq j$.

To complement the numerical test, we can consider the plot of the observed moving sum of residuals $W_n(\mathbf{t})$ versus x_{ij} (or predictor $\mathbf{x}_i^T(\hat{\beta})$) and some realizations from the null distribution. By comparing the observed moving sum of residuals with the realizations, we can detect if there is systematic pattern for the observed moving sum of residuals.

Finally, from a practical perspective, we provide a sequential test procedure to apply the method to real data, though other ways may be used, as follows:

- a. for specified marginal mean, missing data and conditional mean models, do omnibus test to check the departure of the models;
- b. if there is no significant departure of the models, we then use the current models to make inference;
- c. if there is significant departure of the models in (a), check the missing data model π and after repairing π , do the omnibus test again; if there is no significant departure of the models, we then use the current models to make inference;
- d. if there is significant departure of the models with the fixed model π in (c), check the conditional mean model ϕ and after repairing π and ϕ , do the omnibus test; if there is no significant departure of the models, we then use the current models to make inference;
- e. if there is significant departure of the models with the fixed models π and ϕ in (d), check the marginal mean model μ via the omnibus test by repairing π and ϕ .

For the above sequential test procedure, the order of testing these models may change in practice. For example, if departures are detected in (a), one may check for the conditional mean model first then check for the missing data model later. In each step, when checking for the missing data, conditional mean, or marginal mean model, one may employ the method of Lin, Wei and Ying (2002) to check link functions and/or functional forms for covariates.

5. Simulation Studies

In this section, we assess the performance of the proposed methods in terms of type I errors and power. We compare three methods: the lack of fit test for DR estimating equations, the lack of fit test for the weighted estimating equations, and the Su and Wei's method (1991) with complete case. For the lack of fit test for DR estimating equations, we evaluate the following estimating equations: (1) the DR estimating equation with both the missing data and the conditional mean models correctly specified, where we denote as DR-CC; (2) the DR estimating equation with the missing data model correctly specified but the conditional mean model misspecified, where we denote as DR-CI; (3) the DR estimating equation with

the missing data model incorrectly specified but the conditional mean model correctly specified, where we denote as DR-IC; and (4) the DR estimating equation with both the missing data and the conditional mean models incorrectly specified, where we denote as DR-II. For the lack of fit test for the weighted estimating equations, we evaluate the following estimating equations: (5) the weighted estimating equation with correctly specified missing data model, where we denote as WEE-C; and (6) the weighted estimating equation with incorrectly specified missing data model, where we denote as WEE-I. For each study, we choose sample sizes 100, 200, and 500. For each sample size configuration, we consider three missing proportions, 75%, 50% and 25%.

For data generation, we posit that the response Y_i is normally distributed with mean

$$\mu_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2, \quad (13)$$

and standard deviation 1, where x_i is normally distributed with mean 0 and standard deviation 0.5. The true values are $\beta_0 = -1$, $\beta_1 = 1$, and we vary β_2 .

We also assume that an auxiliary variable S_i is also observed, where we assume it is normally distributed with mean $E(S_i|y_i, x_i) = -1 + 0.5y_i + x_i + x_i^2$ and unit variance. It follows that conditional on (s, x) , Y is also normally distributed with a linear form of (s_i, x_i, x_i^2) . To assess the performance of the proposed method, two conditional mean models are considered: $\phi_i(\gamma) = \gamma_0 + \gamma_1 s_i^k + \gamma_2 x_i + \gamma_3 x_i^2$, $k = 1, 2$. So the conditional mean model is posited correctly when $k = 1$ and is incorrect when $k = 2$.

We assume Y_i is subject to missing, and the missing indicator R_i is generated from the following model

$$\text{logit} \pi(s_i, x_i; \alpha) = \alpha_0 + \alpha_1 s_i + \alpha_2 x_i, \quad (14)$$

where we take the true values $\alpha_1 = -1$, $\alpha_2 = 1$, and we vary α_0 to obtain different missing proportions. Similarly, to assess the performance of the proposed method, two missing data models are considered

$$\text{logit} \pi(s_i, x_i; \alpha) = \alpha_0 + \alpha_1 s_i^k + \alpha_2 x_i, \quad k=1, 2. \quad (15)$$

So the missing data model is posited correctly when $k = 1$ and is incorrect when $k = 2$. Figure 1 in the Supplementary Materials is a residual plot of the misspecified conditional mean model ($k = 2$) with a fitted locally-weighted polynomial regression curve when the missing data model is misspecified. This plot can demonstrate how badly misspecified missingness and conditional mean models are when $k = 2$.

In each setting of the simulation studies, we generate 1000 samples, where for each sample, we estimate the p-value under the assumed model through the distribution of G by generating 500 samples z_i , from $N(0, 1)$.

We first compare type I errors of the nominal level 5%, where we assume the response model (13) is correctly specified. Table 1 lists the results. It can be seen that, the type I errors for

methods of DR-CC, DR-CI, DR-IC and WEE-C agree well with the nominal value 5%, especially when the sample size is big, but DR-II, WEE-I and Su & Wei's methods yield type I errors much far from the nominal level 5%.

Next, we compare the power of these methods. We assume the true response model is (13), but we fit a model with mean $\mu_i = \beta_0^* + \beta_1^* x_i$. Table 2 lists the results. The proposed DR test yields more power in DR-CC than DR-IC, DR-CI and DR-II. The proposed DR method yields comparable power between DR-CI and DR-IC, though there is a little bit less power in DR-CI. Furthermore, the proposed DR method yields less power in DR-II than DR-CC, DR-CI and DR-IC. The proposed test for weighted estimating equations yields more power in WEE-C than WEE-I. Su & Wei's method yields less power with smaller sample sizes and more power with bigger sample sizes. The simulation tells us that the power of the proposed method depends on the missing proportion, the value of β_2 and the sample size. As the missing proportion decreases, the power increases; as the absolute value of β_2 increases, the power increases; and as the sample size increases, the power increases. To assess the ability of detecting the misspecification of the missing data model and the conditional mean model, we also calculate the power for the scenario that the mean model is correctly specified, but both the missing data model and the conditional mean model are misspecified, where we denote this model as DR-II* in Table 2. As is seen, in this model setup, the proposed DR test yields smaller power than that in the DR-II model.

We further assess the proposed doubly robust method for checking link functions and functional form of covariates via type I errors and the power. The specification of missing data and conditional mean models are the same as above. For type I errors, we assume the response model (13) is correctly specified, with $\beta_0 = -5$, $\beta_1 = -5$, and we vary β_2 . For the power of checking link functions, we assume the true response model is (13), but we fit a model with a log link function, i.e., we fit model $\log(\mu_i) = \beta_0^* + \beta_1^* x_i$. For the power of checking functional forms of covariates, we assume the true response model is (13), but we fit a model with mean $\mu_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^3$. Table 3 lists the results with sample size 200. The type I errors are all agree with the nominal level 0.05 for DR-CC, DR-CI, and DR-IC when checking for the link function or the functional form of covariates, but the DR-II method yields inflated type I errors. For the power, the proposed DR method yields comparable power among DR-CC, DR-CI and DR-IC when checking the link function (or functional form of covariates), and the proposed DR method yields less power in DR-II.

6. 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 goals of the study is to investigate risk factors that influence the mental status. The response of interest is the Mini-Mental State Examination (MMSE) score (Psychological Assessment Resources, Inc., 2001), which is a screening scale that evaluates orientation to place, orientation to time, registration (immediate repetition of three words), attention and concentration (spelling D-L-R-O-W), recall (recalling the previously repeated three words), language (naming, repetition, reading, writing, comprehension), and visual construction (copy two intersecting pentagons). The MMSE is scored as the number of correctly completed items, with lower scores indicative of

poorer performance and greater cognitive impairment. It ranges from 0 to 30. However, patients may miss the clinic, or refuse to take the test, etc., leading to the incomplete response. Besides the incomplete response, an auxiliary variable, the depression score of patients is observed completely. It is known that the depressive symptoms may be highly correlated with the MMSE. The depression score ranges from 0 to 15. Covariates that may influence the mental status include sex, family history of dementia (fhdm, yes/no), and age. A random sample of 500 subjects who were enrolled in 2005 was used. The missing proportion for the response is about 35%.

Consider the following regression model for the response process

$$\mu_i = \mathbf{X}_i^T \beta, \quad (16)$$

where μ_i is the conditional mean of the MMSE given covariates $\mathbf{X}_i$, and $\mathbf{X}_i$ includes sex, fhdm, and age.

We consider the following conditional mean model for $E(Y_i|S_i, \mathbf{X}_i)$, which is $\gamma_0 + \gamma_1 S_i + \gamma_2^T \mathbf{X}_i$, where S_i denotes the depression score (DEPD).

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

$$\text{logit} \pi_i = \mathbf{u}_i^T \alpha, \quad (17)$$

where $\mathbf{u}_i$ includes sex, fhdm, depression score (DEPD), and age.

We apply the proposed method to check the adequacy of these models. The omnibus test yields a p-value < 0.001 , indicating that there is misspecification of the marginal mean model, or both the missing data model and the conditional mean model. Tables 1 and 2 in the Supplementary Materials and Table 4 list the initial model estimates. As suggested in Section 4.2, we consider checking these models sequentially. We first check the goodness of fit for the missing data model using the method of Lin, Wei and Ying (2002). We start from checking the logistic link function, where the linear predictor in the logistic regression model is a linear combination of sex, fhdm, depression score, and age. The test yields a p-value 0.82, indicating that the logistic link function is appropriate. Next, we check the functional form of age. The test yields a p-value 0.02, indicating that the linear form age may not be appropriate. After checking the data set, we found that the age ranges from 65 to 99 and may not be linearly associated with the log odds. So, instead of using age, we consider the effect of $(\text{age} - 65)^{1/2}$ and redo the test. Now, the test yields a p-value 0.71, indicating that the linear form of $(\text{age} - 65)^{1/2}$ is appropriate in the missing data model. Table 1 in the Supplementary Materials lists the estimates for the missing data model.

We do the omnibus test again with the newly specified missing data model, which yields a p-value < 0.001 . We then check the conditional mean model using the method of Lin, Wei and Ying (2002). We also start from checking the identity link function, where the linear predictor in the linear regression model is a linear combination of sex, fhdm, depression score, and age. The test yields a p-value < 0.001 , indicating that the identity link function is not appropriate. We do a log-transformation for MMSE, $\log(\text{MMSE} + 1)$, and then recheck

the link function, which yields a p-value 0.66, indicating that the log-transformation for MMSE is appropriate. Next, we check the functional form of age by using $\log(\text{MMSE} + 1)$. The test yields a p-value 0.045, indicating that the linear form age may not be appropriate. As suggested in the missing data model, instead of using age, we consider the effect of $(\text{age} - 65)^{1/2}$ and redo the test. Now, the test yields a p-value 0.52, indicating that the linear form of $(\text{age} - 65)^{1/2}$ is appropriate in the conditional mean model. Table 2 in the Supplementary Materials lists the estimate for the conditional mean model and the residual standard error.

We then redo the omnibus test again with the newly specified missing data model and conditional mean model, which yields a p-value < 0.001 . We then consider assessing the goodness of fit for the marginal mean model using the proposed method. We check the identity link function and the functional form of age using the proposed method, where the linear predictor in the regression model is a linear combination of sex, fhdm and age. The test yields a p-value < 0.001 and 0.037 for checking the link function and the functional form of age, respectively, indicating that the identity link function for MMSE and the functional form of age are not appropriate. Figures 2 and 3 in the Supplementary Materials show plots of the observed cumulative sum of residuals (solid lines), with respect to the predictor and age, respectively. Superimposed on the graph are 20 realizations (dotted curves) from the null distribution under the assumption that the doubly robust estimating equations are equal to zero. It appears that there are evident systematic trend for the observed cumulative sum of residuals curves, which is consistent to the numerical assessment. Similarly as suggested in the missing data model and conditional mean model, we consider the effect of $(\text{age} - 65)^{1/2}$, and for the MMSE, we do a log-transformation $\log(\text{MMSE} + 1)$. We refit the model using $(\text{age} - 65)^{1/2}$ and $\log(\text{MMSE} + 1)$ instead of age and MMSE. Now the omnibus test yields a p-value 0.47, indicating that the identity link function for $\log(\text{MMSE} + 1)$ with a linear combination of sex, fhdm and $(\text{age} - 65)^{1/2}$ is appropriate. To complement the test result, we also plot the cumulative sum of residuals to assess the adequacy of the model in Figures 1 and 2 by using $(\text{age} - 65)^{1/2}$ and $\log(\text{MMSE} + 1)$ instead of age and MMSE. It appears that there are no systematic trend for the observed cumulative sum of residuals curves, which is consistent to the numerical assessment.

Table 4 lists the initial and final model estimates for doubly robust estimate, WEE and complete case GEE analyses. The residual standard error is the complete case standard error of $\hat{\theta}_{1j}$. It can be seen that the final model yields different results from the initial model. Sex is not significant in the initial model but significant in the final model. For the final model, the doubly robust estimate gives comparable estimates with the WEE method, but they are different from the complete case GEE analyses, which makes sense since the missing data model may be appropriate, and the missing completely at random assumption is not true (this can be seen from the estimates of the missing data model, the coefficient of DEPD is significant). As is expected, the doubly robust method yields a more precise estimate (suggesting greater efficiency) than the WEE method. The doubly robust and WEE methods reveal that sex has a significant effect on the mental status, but family history of dementia and age have no significant effect on the mental status.

7. Conclusion and Discussion

In this paper, we have proposed a tool to test reliability of the IPWEE and the doubly robust estimates, which is a global goodness-of-fit test of the marginal mean of response, missing data model, or conditional mean model. If departures are detected, misspecification of models may be from the marginal mean of response, or both the missing data model and the conditional mean model. In this situation, we suggest checking these models sequentially as suggested in Section 4. Simulation studies demonstrate that the proposed method works well in terms of the type I errors and power.

The proposed method is more powerful at detecting departures in the marginal mean model of interest than in detecting departures in the missing data and conditional mean models as the estimating equation (6) is constructed specifically to be informative for β , not to be informative for either γ or α . So, the proposed method is more powerful to detect departures in specification of the marginal mean model, in possible presence of misspecified missing data model or the conditional mean model. If our method leads to a rejection of the main regression model, one has to employ the methods by Su and Wei (1991) and Lin, Wei and Ying (2002) to check the correctness of the missing data model and the conditional mean model before one can really reject the main regression model.

The doubly robust method is mainly applied to the continuous outcome, usually normally distributed. In general, it is difficult to extend the doubly robust method to non-normal data, as it is usually difficult to choose model of form (2) to be compatible with the model in equation (1) (Carpenter et al., 2006). As far as we know, little work devoted to the doubly robust method for discrete outcomes due to this reason. Similarly, the proposed method may not be easily applied to discrete outcomes. A future research direction is to develop doubly robust methods for non-normal data and extend our proposed method to this situation following the idea of Clayton et al. (1998). If users do encounter non-normal data, we suggest choosing an appropriate missing data model by using the goodness of fit test so that the doubly robust method is still valid no matter the model of form (2) is compatible with the model in equation (1) or not.

We focussed here primarily on the lack of fit test regarding the response that is subject to missing. The covariates subject to missing are also very common in practice. A future research is to extend this method to the missing outcome and/or missing covariates problem. The idea is similar to the proposed method but the techniques are a little different. This is still under investigation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the Editor, associated editor and two referees for their constructive comments and suggestions, which have improved this paper greatly. Dr. Chen was supported in part by National Institute on Aging grant U01AG016976.

References

- Baker SG, Izmirlian G, Kipnis V. Resolving paradoxes involving surrogate endpoints. *Journal of the Royal Statistical Society, Series A*. 2005; 168:753–762.
- Baker SG. Surrogate endpoints: wishful thinking or reality? (Editorial). *Journal of the National Cancer Institute*. 2006; 98(8):502–503. [PubMed: 16622112]
- Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics*. 2005; 61:962–972. [PubMed: 16401269]
- 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.
- Burzykowski, T.; Molenberghs, G.; Buyse, M. *The Evaluation of Surrogate Endpoints*. New York: Springer; 2005.
- 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]
- 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.
- 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]
- Clayton D, Spiegelhalter D, Dunn G, Pickles A. Analysis of longitudinal binary data from multiphase sampling. *Journal of the Royal Statistical Society, Series B*. 1998; 60:71–87.
- 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]
- 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.
- Duan X, Qin J, Wang Q. Optimal estimation in surrogate outcome regression problems. *Canadian Journal of Statistics*. 2010; 38:1–14.
- Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Annals of Internal Medicine*. 1996; 125:605–613. [PubMed: 8815760]
- Hosmer DW, Lid Hjort N. Goodness-of-fit processes for logistic regression: simulation results. *Statistics in Medicine*. 2002; 21:2723–2738. [PubMed: 12228887]
- Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*. 1952; 47:663–685.
- 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.
- Leung D. Statistical methods in surrogate endpoints. *Journal of the Royal Statistical Society, Series A*. 2001; 164:485–503.
- Lin DY, Wei LJ, Ying Z. Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika*. 1993; 80:557–572.
- Lin DY, Wei LJ, Ying Z. Model-checking techniques based on cumulative residuals. *Biometrics*. 2002; 58:1–12. [PubMed: 11890304]
- Little, RJA.; Rubin, DB. *Statistical Analysis with Missing Data*. 2nd ed.. John Wiley and Sons, Inc.; 2002.
- 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]
- Pan Z, Lin DY. Goodness-of-fit methods for generalized linear mixed models. *Biometrics*. 2002; 61:1000–1009. [PubMed: 16401273]
- Psychological Assessment Resources, Inc.. 16204 North Florida Avenue, Lutz, FL 33549, from the Mini Mental LLC, Inc. Psychological Assessment Resources, Inc; 2001. Published by
- 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.

- 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.
- Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable dropout using semiparametric nonresponse models (with discussion). *Journal of the American Statistical Association*. 1999; 94:1096–1120.
- Seaman S, Copas A. Doubly robust generalized estimating equations for longitudinal data. *Statistics in Medicine*. 2009; 28:937–955. [PubMed: 19153970]
- Su JQ, Wei LJ. A lack-of-fit test for the mean function in a generalized linear model. *Journal of the American Statistical Association*. 1991; 86:420–426.
- Tsiatis, AA. *Semiparametric Theory and Missing Data*. New York: Springer-Verlag; 2006.
- Tsiatis AA, Davidian M, Cao W. Improved doubly robust estimation when data are monotonely coarsened, with application to longitudinal studies with dropout. *Biometrics*. 2010
- Van der Laan, MJ.; Robins, JM. *Unified Methods for Censored Longitudinal Data and Causality*. New York: Springer; 2003.
- Wittes J, Lakatos E, Probstfield J. Surrogate endpoints in clinical trials: Cardiovascular disease. *Statistics in Medicine*. 1989; 8:415–425. [PubMed: 2727465]

Figure 1.

Plot of observed moving sum of residuals versus predictors (solid line) and 20 simulated realizations (dotted curves) from the null distribution assuming the doubly robust estimating equations are equal to zero, and $\log(\text{MMSE}+1)$ has a linear form of sex, fhdem and $(\text{age} - 65)^{1/2}$ for the marginal mean model

Figure 2.

Plot of observed moving sum of residuals versus age (solid line) and 20 simulated realizations (dotted curves) from the null distribution assuming the doubly robust estimating equations are equal to zero, and $\log(\text{MMSE} + 1)$ has a linear form of sex, fhdm and $(\text{age} - 65)^{1/2}$ for the marginal mean model

Table 1

Type I errors for the nominal value 0.05 for the omnibus test: the missing proportions are about 75%, 50% and 25% with $\alpha_0 = -2.5, -0.5$ and 0, respectively; DR-CC: correct missing data and conditional mean models; DR-CI: correct missing data but incorrect conditional mean models; DR-IC: correct conditional mean but incorrect missing data models; DR-II: incorrect missing data and conditional mean models; WEE-C: weighted estimating equation (WEE) method with correct missing data model; WEE-I: WEE with incorrect missing data model; Su & Wei: Su and Wei's method with complete case

n	α_0	β_2	DR-CC	DR-CI	DR-IC	DR-II	WEE-C	WEE-I	Su & Wei
100	-2.5	0.8	0.060	0.060	0.055	0.122	0.060	0.107	0.120
100	-2.5	1	0.056	0.060	0.055	0.121	0.055	0.115	0.120
100	-0.5	0.8	0.048	0.060	0.055	0.135	0.060	0.105	0.108
100	-0.5	1	0.055	0.055	0.056	0.127	0.058	0.107	0.108
100	0	0.8	0.040	0.040	0.045	0.137	0.060	0.090	0.098
100	0	1	0.060	0.060	0.056	0.130	0.050	0.090	0.078
200	-2.5	0.8	0.055	0.057	0.050	0.123	0.055	0.105	0.122
200	-2.5	1	0.055	0.054	0.045	0.121	0.055	0.112	0.122
200	-0.5	0.8	0.048	0.045	0.055	0.134	0.047	0.095	0.098
200	-0.5	1	0.045	0.053	0.050	0.131	0.055	0.095	0.098
200	0	0.8	0.047	0.052	0.055	0.131	0.060	0.120	0.106
200	0	1	0.050	0.050	0.052	0.130	0.054	0.105	0.107
500	-2.5	0.8	0.050	0.050	0.050	0.120	0.054	0.119	0.125
500	-2.5	1	0.056	0.055	0.060	0.122	0.055	0.109	0.126
500	-0.5	0.8	0.050	0.060	0.054	0.135	0.057	0.115	0.119
500	-0.5	1	0.054	0.055	0.055	0.133	0.059	0.135	0.119
500	0	0.8	0.050	0.053	0.050	0.131	0.054	0.115	0.118
500	0	1	0.050	0.050	0.050	0.130	0.053	0.115	0.117

Table 2

Power for the omnibus test: the missing proportions are about 75%, 50% and 25% with $\alpha_0 = -2.5$, -0.5 and 0 , respectively; DR-CC: incorrect mean model, correct missing data and conditional mean models; DR-CI: incorrect mean model, correct missing data but incorrect conditional mean models; DR-IC: incorrect mean model, correct conditional mean but incorrect missing data models; DR-II: incorrect mean model, incorrect missing data and conditional mean models; WEE-C: incorrect mean model, weighted estimating equation (WEE) method with correct missing data model; WEE-I: WEE with incorrect missing data model; Su & Wei: incorrect mean model, Su and Wei's method with complete case; DR-II*: correct mean model, incorrect missing data and conditional mean models

n	α_0	β_2	DR-CC	DR-CI	DR-IC	DR-II	WEE-C	WEE-I	Su & Wei	DR-II*
100	-2.5	0.8	0.766	0.709	0.737	0.687	0.761	0.359	0.396	0.554
100	-2.5	1	0.798	0.745	0.779	0.668	0.794	0.415	0.414	0.661
100	-0.5	0.8	0.814	0.714	0.744	0.696	0.778	0.370	0.426	0.614
100	-0.5	1	0.860	0.815	0.830	0.719	0.836	0.451	0.451	0.707
100	0	0.8	0.836	0.759	0.784	0.724	0.806	0.510	0.444	0.626
100	0	1	0.874	0.823	0.831	0.766	0.872	0.532	0.442	0.720
200	-2.5	0.8	0.881	0.776	0.792	0.776	0.865	0.376	0.873	0.588
200	-2.5	1	0.922	0.836	0.852	0.751	0.873	0.377	0.804	0.628
200	-0.5	0.8	0.919	0.844	0.851	0.778	0.868	0.444	0.861	0.672
200	-0.5	1	0.961	0.884	0.909	0.799	0.913	0.470	0.838	0.734
200	0	0.8	0.928	0.888	0.904	0.802	0.916	0.577	0.812	0.712
200	0	1	0.970	0.920	0.938	0.811	0.943	0.683	0.827	0.797
500	-2.5	0.8	0.970	0.869	0.884	0.867	0.932	0.394	0.910	0.724
500	-2.5	1	0.994	0.852	0.896	0.895	0.967	0.422	0.956	0.736
500	-0.5	0.8	0.997	0.948	0.956	0.913	0.982	0.570	0.930	0.860
500	-0.5	1	1.000	0.958	0.958	0.932	0.984	0.598	0.935	0.878
500	0	0.8	0.998	0.969	0.978	0.924	0.992	0.778	0.934	0.908
500	0	1	1.000	0.982	0.984	0.954	0.992	0.822	0.950	0.917

Table 3

Type I errors for the nominal value 0.05 and power for the link and functional form of covariate test: the missing proportions are about 75%, 50% and 25% with $\alpha_0 = -2.5, -0.5$ and 0, respectively; DR-CC: correct missing data and conditional mean models; DR-CI: correct missing data but incorrect conditional mean models; DR-IC: correct conditional mean but incorrect missing data models; DR-II: incorrect missing data and conditional mean models

<i>n</i>	α_0	β_2	Type I errors				Power			
			DR-CC	DR-CI	DR-IC	DR-II	DR-CC	DR-CI	DR-IC	DR-II
			Check for link function							
200	-2.5	10	0.048	0.045	0.044	0.505	0.845	0.838	0.840	0.840
200	-2.5	5	0.046	0.048	0.045	0.180	0.822	0.816	0.835	0.825
200	-0.5	10	0.044	0.050	0.047	0.310	0.927	0.910	0.925	0.905
200	-0.5	5	0.047	0.051	0.052	0.130	0.918	0.917	0.917	0.890
200	0	10	0.051	0.047	0.044	0.280	0.931	0.911	0.929	0.923
200	0	5	0.049	0.049	0.047	0.125	0.919	0.918	0.917	0.914
Check for functional form of x^2										
200	-2.5	10	0.049	0.046	0.044	0.175	0.988	0.971	0.970	0.868
200	-2.5	5	0.049	0.045	0.053	0.132	0.924	0.926	0.919	0.826
200	-0.5	10	0.047	0.049	0.048	0.130	0.995	0.983	0.981	0.886
200	-0.5	5	0.045	0.052	0.053	0.112	0.951	0.951	0.942	0.867
200	0	10	0.052	0.051	0.046	0.105	0.995	0.984	0.981	0.911
200	0	5	0.047	0.049	0.047	0.101	0.953	0.952	0.946	0.894

Table 4

Point estimate for the National Alzheimer’s Coordinating Center Uniform Dataset: marginal mean model for MMSE (the final model is applied to the $\log(\text{MMSE} + 1)$)

Parameter	Doubly Robust			Weighted EE			GEE		
	Est	StdErr	p	Est	StdErr	p	Est	StdErr	p
Initial model									
Intercept	29.007	4.015	< 0.001	31.009	5.656	< 0.001	27.948	5.614	< 0.001
sex	−0.922	0.702	0.190	−0.716	0.977	0.464	−0.513	0.978	0.601
fhdem	0.054	0.759	0.943	0.048	0.705	0.946	0.077	1.027	0.940
Age	−0.070	0.051	0.171	−0.120	0.071	0.095	−0.084	0.071	0.237
Final model									
Intercept	3.101	0.084	< 0.001	3.110	0.102	< 0.001	2.925	0.157	< 0.001
sex	−0.181	0.072	0.012	−0.183	0.079	0.021	−0.028	0.118	0.808
fhdem	0.045	0.067	0.503	0.057	0.082	0.485	0.033	0.123	0.788
(age − 65) ^{1/2}	−0.001	0.004	0.602	0.000	0.005	0.981	−0.001	0.008	0.823
Resid. stdErr.	0.605								