 DOI: 10.1111/biom.12159

Semiparametric Approach for Non-Monotone Missing Covariates in a Parametric Regression Model

Samiran Sinha,[1] Krishna K. Saha,[2] and Suojin Wang[1,*]

[1]Department of Statistics, Texas A&M University, College Station, Texas 77843, U.S.A.
[2]Department of Mathematical Sciences, Central Connecticut State University, New Britain, Connecticut 06050, U.S.A.
email: sjwang@stat.tamu.edu

Summary. Missing covariate data often arise in biomedical studies, and analysis of such data that ignores subjects with incomplete information may lead to inefficient and possibly biased estimates. A great deal of attention has been paid to handling a single missing covariate or a monotone pattern of missing data when the missingness mechanism is missing at random. In this article, we propose a semiparametric method for handling non-monotone patterns of missing data. The proposed method relies on the assumption that the missingness mechanism of a variable does not depend on the missing variable itself but may depend on the other missing variables. This mechanism is somewhat less general than the completely non-ignorable mechanism but is sometimes more flexible than the missing at random mechanism where the missingness mechansim is allowed to depend only on the completely observed variables. The proposed approach is robust to misspecification of the distribution of the missing covariates, and the proposed mechanism helps to nullify (or reduce) the problems due to non-identifiability that result from the non-ignorable missingness mechanism. The asymptotic properties of the proposed estimator are derived. Finite sample performance is assessed through simulation studies. Finally, for the purpose of illustration we analyze an endometrial cancer dataset and a hip fracture dataset.

Key words: Dimension reduction; Estimating equations; Missing at random; Non-ignorable missing data; Robust method.

1. Introduction

In observational biomedical studies or survey designs, we often encounter partially missing variables, or variables that are not observed or recorded for all the subjects in a study. For example, in the first motivating dataset from the Los Angeles Endometrial Cancer Study, three of the five covariates are partially missing, and in the second dataset on hip fractures, all nine covariates are partially missing. The simplest method, commonly known as complete case (CC) analysis, ignores the subjects with partially missing entries and takes into account the subjects only with complete information on all covariates. This approach may lose efficiency. It may also lead to biased estimates depending on the missingness mechanism. The problem can be severe especially when there are multiple covariates with missing values, even if each of them contains a low to moderate percentage of missing entries. The aims of this article are to introduce a new missingness mechanism for non-monotone missing data and to propose a semiparametric method to estimate the model parameters.

Suppose that our interest is in a finite dimensional parameter β in the parametric regression model $f(Y|X, Z; \beta)$ where Y is the outcome variable and $X = (X_1, \ldots, X_p)^{\mathrm{T}}$ and $Z = (Z_1, \ldots, Z_q)^{\mathrm{T}}$ are the explanatory variables. We assume that Y and Z are observed for all subjects, whereas all the components of X are partially observed. Let (X_i, Y_i, Z_i), $i = 1, \ldots, n$, be independent and identically distributed (iid) copies of (X, Y, Z), where n is the sample size. Define the missingness indicator variables

$$R_{ij} \equiv \begin{cases} 1 & \text{if } X_{ij} \text{ is observed,} \\ 0 & \text{if } X_{ij} \text{ is not observed,} \end{cases} \quad j = 1, \ldots, p$$

and $X_{i(-r)} \equiv (X_{i1}, \ldots, X_{i(r-1)}, X_{i(r+1)}, \ldots, X_{ip})^{\mathrm{T}}$. There are two possibilities for multiple missing covariates: (1) monotone missing data and (2) non-monotone missing data. For monotone missing data, if X_{ij} is missing, then X_{ik}, $k > j$ are all missing, that is, $\mathrm{pr}(R_{i(j+1)} = \cdots = R_{ip} = 0 | R_{ij} = 0) = 1$ for any $j = 1, \ldots, (p-1)$. In this article, we focus on non-monotone missing covariate data.

Regarding the missing data mechanism, the missing at random (MAR) assumption is widely used, under which the probability that $R_{ij} = 1$ depends only on the observed quantities (Rubin, 1976). Under the MAR mechanism and parameter distinctness assumption, valid likelihood based inferences may be carried out without using a model for the missingness mechanism. For handling MAR covariate data, multiple imputation is a commonly used method. Within the broader context of imputation approach, Reilly and Pepe (1995) proposed the mean-score approach and Chatterjee, Chen, and Breslow (2003) proposed a pseudo-score approach which is usually more efficient than the mean-score approach if the assumed model for the missingness mechanism is correct. Robins, Rotnitzky, and Zhao (1994) proposed an efficient method within the class of inverse probability weighted (IPW) estimating

equations for handling the MAR data which is similar in spirit to the Horvitz–Thompson estimator. Lipsitz, Ibrahim, and Zhao (1999) proposed a doubly robust method that blends the likelihood based approach and the weighted estimating equations. Ibrahim, Chen, and Lipstiz (1999) proposed a full likelihood based method using the EM algorithm. They assumed a parametric model for the partially missing covariate. Under the MAR assumption, Chen (2004) proposed a semiparametric method for handling multiple missing covariates for any arbitrary pattern of missing data. He assumed a parametric model for the odds ratio between any two missing covariates, which is then used in the likelihood formation.

Non-ignorable (NI) mechanism happens if the probability that $R_{ij} = 1$ depends on values of the completely observed and partially observed variables, meaning that the missingness mechanism of X_{ij} may depend on all components of $X_i = (X_{i1}, \ldots, X_{ip})^{\mathrm{T}}$ along with Y_i and Z_i. For NI missing data, the model parameters are estimated by jointly maximizing the likelihood of the data generation process and the missingness process. This requires a parametric model specification of the distribution of the missing covariates and the missingness process. Consistent parameter estimates are obtained under the assumption that both the models are correctly specified. Ibrahim, Lipsitz, and Chen (1999) wrote a comprehensive review of NI missing covariate data scenario, and they considered arbitrary missing data patterns. One disadvantage of the NI missingness mechanism is that data provide little or no information regarding the parameters in the missingness models. Hence, model parameters are weakly identifiable (Little, 1995).

Non-monotone missing data are usually handled by assuming the MAR mechanism (Robins and Gill, 1997) or the NI mechanism (Robins, 1997). In the former case, analysis can be done without assuming any model for the missingness mechanism and any parametric model for the partially missing covariates. In the latter case, analysis is usually done by positing a parametric model for the missingness mechanism and partially missing covariates. Except for some special cases (e.g., Lei and Wang, 2001), the MAR and NI assumptions are generally not testable.

To fill the substantial gap between the MAR and NI mechanisms, we introduce another mechanism which is somewhat less general than the NI mechanism but facilitates our calculations as it can be used to tackle the identifiability issue. To handle the non-monotone pattern of missing data, we define missingness mechanism "NI-" when the probability that $R_{ij} = 1$ may depend on all variables except the jth covariate X_{ij}, that is, $\mathrm{pr}(R_{ij} = 1|Y_i, X_i, Z_i) = \mathrm{pr}(R_{ij} = 1|Y_i, X_{i(-j)}, Z_i)$. Thus, we assume that R_{ij} may depend on the other variables which are completely observed or partially missing, but does not depend on the values of the jth covariate X_{ij}. Further, we assume that under the NI-mechanism, (1) $R_{i1}, \ldots, R_{ip}$ are independent conditional on X_i, Y_i, Z_i, and (2) for every j, $\mathrm{pr}(R_{ij} = 1|Y_i, X_i, Z_i) > 0$ with probability one for every combination of (Y_i, X_i, Z_i). Under the independence assumption (1), the NI-mechanism helps to reduce the identifiability issue encountered by the non-ignorable missing data. Consequently, we do not need to make a parametric model assumption regarding the distribution of the missing covariates. Our NI-mechanism is partly non-ignorable because the probability model for R_j may depend on $(X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_p)$ which are partially observed. In summary, the proposed NI-mechanism is sometimes more flexible than the MAR mechanism where the missingness mechanism is allowed to depend only on the completely observed variables. Further, the NI-mechanism allows us to estimate the model parameters without specifying a parametric model for the missing covariate that is needed for the NI missingness mechanism. When the MAR assumption is in doubt, it may be preferable to assume the NI-mechanism and use the proposed method of estimation.

The remainder of this article is outlined as follows: Sections 2 and 3 contain the detailed methodology and asymptotic properties for the case of two partially missing covariates. An extension of the method for more than two partially missing covariates is discussed in Section 4. Sections 5 and 6 contain simulation studies and two real data examples, respectively. Some concluding remarks are given in Section 7. Technical material is provided in the Appendix.

2. Estimation Methodology

2.1. *Background*

The goal is to estimate β. Without any missing entries in X, and under certain regularity conditions, the MLE of β is obtained by solving the score function $S_{\mathrm{nm}} = \sum_{i=1}^{n} S_\beta(Y_i, X_i, Z_i) = 0$, where $S_\beta(Y, X, Z) = \partial \log\{f(Y|X, Z; \beta)\}/\partial\beta$ for $f(Y|X, Z; \beta)$ satisfying some standard regularity conditions. For the CC analysis, β is estimated by solving $S_{\mathrm{cc}} = \sum_{i=1}^{n} (\prod_{j=1}^{p} R_{ij}) S_\beta(Y_i, X_i, Z_i) = 0$. Note that Reilly and Pepe (1995) and Chatterjee et al. (2003) both considered the case where all R_{ij}, $j = 1, \ldots, p$ are equal for a given i, that is, either $R_{i1} = \cdots = R_{ip} = 1$ or $R_{i1} = \cdots = R_{ip} = 0$ and their missingness mechanism depends only on the observable quantities.

2.2. *Proposed Method for NI-Missing Data*

In this Section, we consider the case of $p = 2$. We will use the following notation $\pi_{ij}(Y_i, X_{i(-j)}, Z_i) = 1 - \bar{\pi}_{ij}(Y_i, X_{i(-j)}, Z_i) \equiv \mathrm{pr}(R_{ij} = 1|Y_i, X_{i(-j)}, Z_i)$. We first assume that both the covariates X and Z are discrete with fixed numbers of categories while Y can be either discrete or continuous. In the following derivation, any integration with respect to a discrete covariate implies a summation. The likelihood of the data is

$$
\begin{aligned}
L = & \prod_{i: R_{i1} = R_{i2} = 1} \pi_{i1}(Y_i, X_{i(-1)}, Z_i) \pi_{i2}(Y_i, X_{i(-2)}, Z_i) f(Y_i|X_i, Z_i) \\
& \times f(X_{i1}, X_{i2}|Z_i) \prod_{i: R_{i1} = 1, R_{i2} = 0} \left\{ \int \pi_{i1}(Y_i, X_{i(-1)}, Z_i) \right. \\
& \left. \times f(Y_i|X_i, Z_i) f(X_{i1}, X_{i2}|Z_i)\, \mathrm{d}X_{i2} \right\} \bar{\pi}_{i2}(Y_i, X_{i(-2)}, Z_i) \\
& \times \prod_{i: R_{i1} = 0, R_{i2} = 1} \bar{\pi}_{i1}(Y_i, X_{i(-1)}, Z_i) \left\{ \int \pi_{i2}(Y_i, X_{i(-2)}, Z_i) \right. \\
& \left. \times f(Y_i|X_i, Z_i) f(X_{i1}, X_{i2}|Z_i)\, \mathrm{d}X_{i1} \right\}
\end{aligned}
$$

$$\times \prod_{i:R_{i1}=0,R_{i2}=0} \left\{ \int \bar{\pi}_{i1}(Y_i, X_{i(-1)}, Z_i)\bar{\pi}_{i2}(Y_i, X_{i(-2)}, Z_i) \times f(Y_i|X_i, Z_i) f(X_{i1}, X_{i2}|Z_i)\, \mathrm{d}X_{i1}\, \mathrm{d}X_{i2} \right\}.$$

The score function for estimating β is

$$\begin{aligned} S_{f_x,\beta} = & \sum_{i:R_{i1}=R_{i2}=1} S_\beta(Y_i, X_i, Z_i) \\ & + \sum_{i:R_{i1}=1,R_{i2}=0} \int S_\beta(Y_i, X_i, Z_i) q_{i,2}(\pi_{i1})\, \mathrm{d}X_{i2} \\ & + \sum_{i:R_{i1}=0,R_{i2}=1} \int S_\beta(Y_i, X_i, Z_i) q_{i,1}(\pi_{i2})\, \mathrm{d}X_{i1} \\ & + \sum_{i:R_{i1}=0,R_{i2}=0} \int S_\beta(Y_i, X_i, Z_i) q_{i,12}(\bar{\pi}_{i1}\bar{\pi}_{i2})\, \mathrm{d}X_i, \quad (1) \end{aligned}$$

where for $r = 1, 2, 12$, $q_{i,r}(\omega) = f(Y_i|X_i, Z_i)\omega(Y_i, X_i, Z_i) f(X_{ir}| X_{i(-r)}, Z_i)/\int f(Y_i|X_i, Z_i)\omega(Y_i, X_i, Z_i)\ f(X_{ir}|X_{i(-r)}, Z_i)\, \mathrm{d}X_{ir}$ is a function of the given parametric model $f(Y_i|X_i, Z_i)$, the distribution of the missing covariate given the completely observed covariates, and ω. Here $r = 12$ means both X_{i1} and X_{i2} are missing. For obtaining $q_{i,2}(\pi_{i1})$, $q_{i,1}(\pi_{i2})$, and $q_{i,12}(\bar{\pi}_{i1}\bar{\pi}_{i2})$, we set $\omega = \pi_{i1}, \pi_{i2}$, and $\bar{\pi}_{i1}\bar{\pi}_{i2}$, respectively.

For estimating $q_{i,r}(\omega)$ we face two issues: estimation of the missing covariate distribution and of the missingness probabilities (selection probabilities). One obvious approach is to model the missingness probability and the covariate distribution parametrically. However, modeling the distribution of the covariates parametrically is not an easy task, and model misspecification can lead to biased estimates. Therefore, we use an empirical approach to estimate $f(X_{ir}|X_{i(-r)}, Z_i)$ and use a parametric model for the missingness probabilities. Suppose that the missingness probabilities are identified by a finite dimensional parameter $\alpha = (\alpha_1^T, \alpha_2^T)^{\mathrm{T}}$, where α_k is for X_k for $k = 1, 2$. We will denote $\omega(Y_i, X_i, Z_i)$ by $\omega(Y_i, X_i, Z_i, \alpha)$. Adopting the technique of Chatterjee et al. (2003) to our setting, we write the conditional density of X_{ir} given $X_{i(-r)}$ and Z_i in terms of the conditional density of X_{ir} given $X_{i(-r)}$ and Z_i in the completely observed data $V = \{i : R_{i1} = R_{i2} = 1\}$ as

$$f(X_{ir}|X_{i(-r)}, Z_i) = \frac{f(X_{ir}|X_{i(-r)}, Z_i, R_{i1}=R_{i2}=1)\mathrm{pr}(R_{i1}=R_{i2}=1|X_{i(-r)}, Z_i)}{\mathrm{pr}(R_{i1} = R_{i2} = 1|X_i, Z_i)}. \quad (2)$$

Plugging this into the expression of $q_{i,r}$ we obtain $q_{i,r}(\omega) = h(Y_i, X_i, Z_i; \omega) f(X_{ir}|X_{i(-r)}, Z_i,\ R_{i1} = R_{i2} = 1)\,/\int h(Y_i,\ X_i,\ Z_i; \omega) f(X_{ir}|X_{i(-r)}, Z_i, R_{i1} = R_{i2} = 1)\, \mathrm{d}X_{ir}$, where $h(X_i, Y_i, Z_i; \omega) = f(Y_i|X_i, Z_i)\omega(X_i, Y_i,\ \ Z_i)/\mathrm{pr}(R_{i1} = R_{i2} = 1|X_i, Z_i)$. Following a strategy similar to that of Chatterjee et al. (2003) we obtain $\mathrm{pr}(R_{i1} = R_{i2} = 1|X_i, Z_i) = \int \mathrm{pr}(R_{i1} = R_{i2} = 1|X_i, Y_i, Z_i) f(Y|X_i, Z_i; \beta)\, \mathrm{d}Y$.

Therefore, $q_{i,1}(\pi_{i2})$, $q_{i,2}(\pi_{i1})$, and $q_{i,12}(\bar{\pi}_{i1}\bar{\pi}_{i2})$ involve $f(X_{i1}|X_{i(-1)}, Z_i, R_{i1} = R_{i2} = 1)$, $f(X_{i2}|X_{i(-2)},\ Z_i, R_{i1} = R_{i2} = 1)$, and $f(X_i|Z_i, R_{i1} = R_{i2} = 1)$, respectively. Note that these functions must be estimated empirically based on the completely observed data V. Observe that while $\mathrm{pr}(R_{i1} = R_{i2} = 1|X_{i(-r)}, Z_i)$ appears in the numerator of (2), it does not appear in $q_{i,r}(\omega)$. An estimator of $q_{i,r}(\omega)$ denoted by $\widehat{q}_{i,r}(\omega)$ is obtained after replacing $f(X_{ir}|X_{i(-r)}, Z_i, R_{i1} = R_{i2} = 1)$ by its empirical estimates $\widehat{f}(X_{ir}|X_{i(-r)}, Z_i, R_{i1} = R_{i2} = 1) = \mathrm{d}\widehat{F}(X_{ir}|X_{i(-r)}, Z_i, R_{i1} = R_{i2} = 1)$, where

$$\widehat{F}(x|X_{i(-r)}, Z_i, R_{i1} = R_{i2} = 1) = \frac{\sum_{j=1}^n I(X_{jr} \le x, X_{j(-r)} = X_{i(-r)}, Z_j = Z_i, R_{j1} = R_{j2} = 1)}{\sum_{j=1}^n I(X_{j(-r)} = X_{i(-r)}, Z_j = Z_i, R_{j1} = R_{j2} = 1)}$$

and in particular when $r = 12$,

$$\widehat{F}(x_1, x_2|Z_i, R_{i1} = R_{i2} = 1) = \frac{\sum_{j=1}^n I(X_{j1} \le x_1, X_{j2} \le x_2, Z_j = Z_i, R_{j1} = R_{j2} = 1)}{\sum_{j=1}^n I(Z_j = Z_i, R_{j1} = R_{j2} = 1)}.$$

After replacing $q_{i,r}$ by $\widehat{q}_{i,r}$ $(r = 1, 2, 12)$ in (1) we obtain the estimated score functions

$$\begin{aligned} S_{\widehat{f}_x,\beta} = & \sum_{i:R_{i1}=R_{i2}=1} S_\beta(Y_i, X_i, Z_i) \\ & + \sum_{i:R_{i1}=1,R_{i2}=0} \int S_\beta(Y_i, X_i, Z_i)\widehat{q}_{i,2}(\pi_{i1})\, \mathrm{d}X_{i2} \\ & + \sum_{i:R_{i1}=0,R_{i2}=1} \int S_\beta(Y_i, X_i, Z_i)\widehat{q}_{i,1}(\pi_{i2})\, \mathrm{d}X_{i1} \\ & + \sum_{i:R_{i1}=0,R_{i2}=0} \int S_\beta(Y_i, X_i, Z_i)\widehat{q}_{i,12}(\bar{\pi}_{i1}\bar{\pi}_{i2})\, \mathrm{d}X_i. \end{aligned}$$

For estimating β we also need to estimate α_1 and α_2. Define $S_{\alpha_k}(R_{ik}, Y_i, X_i, Z_i) = R_{ik}\partial\log\{\pi_{ik}(Y_i,\ X_{i(-k)}, Z_i)\}/\partial\alpha_k + (1 - R_{ik})\partial\log\{\bar{\pi}_{ik}(Y_i, X_{i(-k)}, Z_i)\}/\partial\alpha_k$ for $k = 1, 2$. Then we can estimate α_1 and α_2 by solving the estimating equations

$$\begin{aligned} S_{\widehat{f}_x,\alpha_1} = & \sum_{i:R_{i1}=R_{i2}=1} S_{\alpha_1}(R_{i1}, Y_i, X_i, Z_i) \\ & + \sum_{i:R_{i1}=1,R_{i2}=0} \int S_{\alpha_1}(R_{i1}, Y_i, X_i, Z_i)\widehat{q}_{i,2}(\pi_{i1})\, \mathrm{d}X_{i2} \\ & + \sum_{i:R_{i1}=0,R_{i2}=1} S_{\alpha_1}(R_{i1}, Y_i, X_i, Z_i) \\ & + \sum_{i:R_{i1}=0,R_{i2}=0} \int S_{\alpha_1}(R_{i1}, Y_i, X_i, Z_i)\widehat{q}_{i,12}(\bar{\pi}_{i1}\bar{\pi}_{i2})\, \mathrm{d}X_i = 0 \end{aligned}$$

and

$$S_{\widehat{f}_x,\alpha_2} = \sum_{i:R_{i1}=R_{i2}=1} S_{\alpha_2}(R_{i2}, Y_i, X_i, Z_i) + \sum_{i:R_{i1}=0,R_{i2}=1} \int S_{\alpha_2}(R_{i2}, Y_i, X_i, Z_i)\widehat{q}_{i,1}(\pi_{i2})\,\mathrm{d}X_{i1} + \sum_{i:R_{i1}=1,R_{i2}=0} S_{\alpha_2}(R_{i2}, Y_i, X_i, Z_i) + \sum_{i:R_{i1}=0,R_{i2}=0} \int S_{\alpha_2}(R_{i2}, Y_i, X_i, Z_i)\widehat{q}_{i,12}(\bar{\pi}_{i1}\bar{\pi}_{i2})\,\mathrm{d}X_i = 0.$$

If the missingness model of R_r does not depend on $X_{(-r)}$ then the missingness model becomes a model for the MAR data. Therefore, the resulting method will produce consistent estimates when the data are MAR, and the method will reduce to the pseudo-score method of Chatterjee et al. (2003). While there is no direct test to verify this assumption, the NI-mechanism would be worth considering if R_1 (R_2) depends on X_2 (X_1) in the subset $\{i : R_{i2} = 1\}$ ($\{i : R_{i1} = 1\}$).

When any component of X or Z is continuous, a similar method can be developed where conditional probability mass function (pmf) must be replaced by a conditional density function estimated by, say, a kernel method. This approach requires a full scale technical and numerical investigation and is a problem for future research. Alternatively, in this work we approximate the conditional density by the empirical conditional pmf after discretizing the continuous components. Although this method is an approximation, our finite sample numerical studies suggest that it works well. Note that in principle discretization should be refined with sample size so that the conditional pmf converges to the conditional density as n grows to infinity. How to optimize the rate of refinement as a function of the sample size is an interesting problem for future work.

2.3. *The Issue of Identifiability*

The likelihood identifiability is a sophisticated issue in the general missing data context. Through the following simple example we illustrate how the identifiability issue under the NI data is eased under the NI-missing data assumption when the models for the missingness indicators and the covariate distribution are nonparametric.

Suppose that potentially observable data are n iid copies of (Y, X_1, X_2, R_1, R_2) without Z, where Y is a binary response variable, X_1 and X_2 both are scalar binary variables, and R_k is a missingness indicator for X_k for $k = 1, 2$. Let $n_{x_1,x_2,y}$ be the number of observations with $X_1 = x_1$, $X_2 = x_2$, $Y = y$, and for $R_1 = R_2 = 1$; $m_{x_1,-,y}$ be the number of observations with $X_1 = x_1$, missing X_2, and $Y = y$ (i.e., where R_1=1 and $R_2 = 0$); $m_{-,x_2,y}$ be the number of observations with missing X_1, $X_2 = x_2$, and $Y = y$ (i.e., where $R_1 = 0$ and $R_2 = 1$); and $m_{-,-,y}$ be the number of observations with missing X_1 and X_2, and $Y = y$ (i.e., where $R_1 = 0$ and $R_2 = 0$). Thus, $n_{0,0,0} + n_{0,0,1} + n_{0,1,0} + n_{0,1,1} + n_{1,0,0} + n_{1,0,1} + n_{1,1,0} + n_{1,1,1} + m_{-,0,0} + m_{-,0,1} + m_{-,1,0} + m_{-,1,1} + m_{0,-,0} + m_{0,-,1} + m_{1,-,0} + m_{1,-,1} + m_{-,-,0} + m_{-,-,1} = n$. Let $u_{x_1,x_2} = \mathrm{pr}(X_1 = x_1, X_2 = x_2)$, $v_{x_1,x_2} = \mathrm{pr}(Y = 1|X_1 = x_2, X_2 = x_2)$, $\pi^{(1)}_{x_1,x_2,y} = \mathrm{pr}(R_1 = 1|X_1 = x_1, X_2 = x_2, Y = y)$, and $\pi^{(2)}_{x_1,x_2,y} = \mathrm{pr}(R_2 = 1|X_1 = x_1, X_2 = x_2, Y = y)$. The likelihood of the observed data involves 23 parameters ($4v$-parameters, $3u$-parameters, $8\pi^{(1)}$-parameters, and $8\pi^{(2)}$-parameters) whereas the data contain $18 - 1 = 17$ independent patterns; therefore the parameters are not identifiable from the data.

However, this issue may be eased by imposing some model restrictions such as our NI-mechanism. In the NI-mechanism, there are at most $4\pi^{(1)}$-parameters (no X_1 involved) and $4\pi^{(2)}$-parameters (no X_2 involved), resulting in a total of 15 model parameters, which is fewer than the number of observed patterns of the data. Let L be the likelihood of the observed data and $\theta = (u_{00}, u_{01}, u_{10}, v_{00}, v_{01}, v_{10}, v_{11}, \pi^{(1)}_{00}, \pi^{(1)}_{01}, \pi^{(1)}_{10}, \pi^{(1)}_{11}, \pi^{(2)}_{00}, \pi^{(2)}_{01}, \pi^{(2)}_{10}, \pi^{(2)}_{11})^{\mathrm{T}}$ be the set of parameters under the NI-mechanism. In the Supplementary Materials, we show that $E[\partial\log(L)/\partial\theta\{\partial\log(L)/\partial\theta\}^{\mathrm{T}}]$ is non-singular, a sufficient condition for identifiability (Rothenberg, 1971). The issue in a more general setting is interesting and worth investigating. Additionally, partial information regarding α_1 (generally a non-saturated equivalent of $\pi^{(1)}$) involved in $\mathrm{pr}(R_1 = 1|X_2, Y; \alpha_1)$ can be obtained from the data (R_1, X_2, Y) among the subjects with $R_2 = 1$. Note that these subjects' contribution to the likelihood for α_1 is $\prod_{i:R_{i2}=1} \mathrm{pr}^{R_{i1}}(R_{i1} = 1|R_{i2}, X_{i1}, X_{i2}, Y_i)\mathrm{pr}^{1-R_{i1}}(R_{i1} = 0|R_{i2}, X_{i1}, X_{i2}, Y_i) = \prod_{i:R_{i2}=1} \mathrm{pr}^{R_{i1}}(R_{i1} = 1|X_{i2}, Y_i)\mathrm{pr}^{1-R_{i1}}(R_{i1} = 0|X_{i2}, Y_i)$ due to our assumption that $R_1 \perp R_2|X_1, X_2, Y$ and $R_1 \perp X_1|X_2, Y$. To clarify the notation, the R_{i1} inside the probability statements above is a random variable while the R_{i1} on the exponents is the observed value of the random variable. Similarly, information regarding α_2 can be obtained from the data (R_2, X_1, Y) among the subjects with $R_1 = 1$.

2.4. *Computational Steps*

Let $\theta = (\alpha^T, \beta^T)^{\mathrm{T}}$. Then θ is estimated by solving the estimating equations $S_{\widehat{f},\theta} = 0$ iteratively, where $S^T_{\widehat{f}_x,\theta} = (S^T_{\widehat{f}_x,\alpha_1}, S^T_{\widehat{f}_x,\alpha_2}, S^T_{\widehat{f}_x,\beta})$. Let $\alpha_1^{(t)}$, $\alpha_2^{(t)}$, and $\beta^{(t)}$ be the parameter estimates at the tth iteration.

Step 0. Initialize $\alpha_1 = \alpha_1^{(0)}$, $\alpha_2 = \alpha_2^{(0)}$, and $\beta = \beta^{(0)}$.

Step 1. For the $(t + 1)$th iteration first compute $\widehat{q}_{i,r}$ for $r = 1, 2, 12$ using $\alpha_1 = \alpha_1^{(t)}$, $\alpha_2 = \alpha_2^{(t)}$, and $\beta = \beta^{(t)}$. Then insert these quantities into the estimating equations $S_{\widehat{f}_x,\theta} = 0$.

Step 2. Solve for α_1, α_2, and β from $S_{\widehat{f}_x,\theta} = 0$ using the Newton–Raphson method by assuming that $\widehat{q}_{i,r}$ is fixed. Observe that conditional on $\widehat{q}_{i,r}$, $r = 1, 2, 12$, the Newton–Raphson method can be carried out separately for each parameter vector, β, α_1, and α_2, thereby keeping the dimension low.

Step 3. Repeat Steps 1 and 2 until the estimates converge. We denote the resulting estimates by $\widehat{\alpha}$ and $\widehat{\beta}$.

For the Newton–Raphson method of Step 2, we use $\widehat{\theta} = \theta - H_\theta^{-1} S_{\widehat{f},\theta}$, where H_θ is the matrix of the partial derivatives of $S_{\widehat{f},\theta}$ for fixed $\widehat{q}_{i,r}$, $r = 1, 2, 12$. The computational step is similar to the EM algorithm as in every iteration, β, α_1, and

α_2 are determined assuming the conditional distributions $\widehat{q}_{i,r}$, $r = 1, 2, 12$ are known.

3. Asymptotic Properties

We start with introducing some additional notation. Let $S_{\widehat{f},\alpha} = (S^T_{\widehat{f},\alpha_1}, S^T_{\widehat{f},\alpha_2})^{\mathrm{T}}$. For convenience, from now on we will omit the arguments of $\pi_{ij}(Y_i, X_{i(-j)}, Z_i)$, $S_\beta(Y_i, X_i, Z_i)$, and $h(Y_i, X_i, Z_i, \omega)$ and simply use π_{ij}, $S_{i,\beta}$, and $h_i(\omega)$, respectively. Define $D = (D_{ll'})_{ll'}$, where $D_{11} = A_{\alpha\alpha} \equiv \lim_{n\to\infty} n^{-1}\partial S_{\widehat{f},\alpha}/\partial\alpha$, $D_{12} = B \equiv \lim_{n\to\infty} n^{-1}\partial S_{\widehat{f},\alpha}/\partial\beta$, $D_{21} = C \equiv \lim_{n\to\infty} n^{-1}\partial S_{\widehat{f},\beta}/\partial\alpha$, and $D_{22} = A_{\beta\beta} \equiv \lim_{n\to\infty} n^{-1}\partial S_{\widehat{f},\beta}/\partial\beta$, and $S^{\mathrm{adj}}_{i,\widehat{f},\alpha_1} \equiv R_{i1}R_{i2}S_{\alpha_1}(R_{i1}, Y_i, X_i, Z_i) + R_{i1}(1-R_{i2})E\{S_{\alpha_1}(R_{i1}, Y_i, X_i, Z_i)|Y_i, X_{i1}, Z_i, R_{i1} = 1, R_{i2} = 0\} + \Upsilon_{i,\alpha_1,10} + (1-R_{i1})R_{i2}S_{\alpha_1}(R_{i1}, Y_i, X_i, Z_i) + (1-R_{i1})(1-R_{i2})E\{S_{\alpha_1}(R_{i1}, Y_i, X_i, Z_i)|Y_i, Z_i, R_{i1} = R_{i2} = 0\} + \Upsilon_{i,\alpha_1,00}$, $S^{\mathrm{adj}}_{i,\widehat{f},\alpha_2} \equiv R_{i1}R_{i2}S_{\alpha_2}(R_{i2}, Y_i, X_i, Z_i) + R_{i1}(1-R_{i2})S_{\alpha_2}(R_{i2}, Y_i, X_i, Z_i) + (1-R_{i1})R_{i2}E\{S_{\alpha_2}(R_{i2}, Y_i, X_i, Z_i)|Y_i, X_{i2}, Z_i, R_{i1} = 0, R_{i2} = 1\} + \Upsilon_{i,\alpha_2,01} + (1-R_{i1})(1-R_{i2})E\{S_{\alpha_2}(R_{i2}, Y_i, X_i, Z_i)| Y_i, Z_i, R_{i1} = R_{i2} = 0\} + \Upsilon_{i,\alpha_2,00}$, $S^{\mathrm{adj}}_{i,\widehat{f},\beta} \equiv R_{i1}R_{i2}S_\beta(Y_i, X_i, Z_i) + R_{i1}(1-R_{i2})E\{S_\beta(Y_i, X_i, Z_i)|Y_i, X_{i1}, Z_i, R_{i1} = 1, R_{i2} = 0\} + \Upsilon_{i,\beta,10} + (1-R_{i1})R_{i2}E\{S_\beta(Y_i, X_i, Z_i)|Y_i, X_{i2}, Z_i, R_{i1} = 0, R_{i2} = 1\} + \Upsilon_{i,\beta,01} + (1-R_{i1})(1-R_{i2})E\{S_\beta(Y_i, X_i, Z_i)|Y_i, Z_i, R_{i1} = R_{i2} = 0\} + \Upsilon_{i,\beta,00}$, where the expressions for the adjustment terms $\Upsilon_{i,\alpha_1,10}$, $\Upsilon_{i,\alpha_1,00}$, $\Upsilon_{i,\alpha_2,01}$, $\Upsilon_{i,\alpha_2,00}$, $\Upsilon_{i,\beta,10}$, $\Upsilon_{i,\beta,01}$, and $\Upsilon_{i,\beta,00}$ are given in the Appendix. Now we have the following main results.

THEOREM 1. *Under regularity conditions C1–C8 listed in Web Appendix A3, when $n \to \infty$,*
(a) the influence function representation of the estimators is

$$\sqrt{n}\begin{pmatrix}\widehat{\alpha}-\alpha\\ \widehat{\beta}-\beta\end{pmatrix} = -D^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix}S^{\mathrm{adj}}_{i,\widehat{f},\alpha}\\ S^{\mathrm{adj}}_{i,\widehat{f},\beta}\end{pmatrix} + o_p(1);$$

(b) asymptotically $\sqrt{n}(\widehat{\beta}-\beta) \sim \mathrm{Normal}(0, \Sigma_\beta)$, where $\Sigma_\beta = H^{-1}\mathrm{cov}(S^{\mathrm{adj}}_{1,\widehat{f},\beta})H^{-1} - 2H^{-1}\mathrm{cov}(S^{\mathrm{adj}}_{1,\widehat{f},\beta}, S^{\mathrm{adj}}_{1,\widehat{f},\alpha})\ F^{-T}C^TA^{-T}_{\beta\beta} + A^{-1}_{\beta\beta}CF^{-1}\mathrm{cov}(S^{\mathrm{adj}}_{1,\widehat{f},\alpha})F^{-1}C^TA^{-1}_{\beta\beta}$, with $H = A_{\beta\beta} - CA^{-1}_{\alpha\alpha}B$ and $F = A_{\alpha\alpha} - BA^{-1}_{\beta\beta}C$.

The exact expression for D and a sketch proof of the above theorem are given in the Supplementary Materials.

COROLLARY 1. *(a) If we replace $A_{\alpha\alpha}$, $A_{\beta\beta}$, B, and C in Σ_β by $n^{-1}\partial S_{\widehat{f}_x,\alpha}/\partial\alpha$, $n^{-1}\partial S_{\widehat{f}_x,\beta}/\partial\beta$, and $n^{-1}\partial S_{\widehat{f}_x,\alpha}/\partial\beta$, $n^{-1}\partial S_{\widehat{f}_x,\beta}/\partial\alpha$, respectively, then the resulting estimate $\widehat{\Sigma}_\beta$ is consistent for Σ_β;*
(b) $n^{-1}\sum_{i=1}^n S^{\mathrm{adj}}_{i,\widehat{f},\beta}S^{\mathrm{adj}\,T}_{i,\widehat{f},\beta}$, $n^{-1}\sum_{i=1}^n S^{\mathrm{adj}}_{i,\widehat{f},\beta}S^{\mathrm{adj}\,T}_{i,\widehat{f},\alpha}$, and $n^{-1}\sum_{i=1}^n S^{\mathrm{adj}}_{i,\widehat{f},\alpha}S^{\mathrm{adj}\,T}_{i,\widehat{f},\alpha}$ are consistent estimates for $\mathrm{cov}(S^{\mathrm{adj}}_{1,\widehat{f},\beta})$, $\mathrm{cov}(S^{\mathrm{adj}}_{1,\widehat{f},\beta}, S^{\mathrm{adj}}_{1,\widehat{f},\alpha})$, and $\mathrm{cov}(S^{\mathrm{adj}}_{1,\widehat{f},\alpha})$, respectively.

The proof of the corollary is straightforward and thus omitted. Note that the adjustment terms are zero for subjects $i \notin V = \{j : R_{j1} = R_{j2} = 1\}$. For $i \in V$, $\Upsilon_{i,\alpha_1,10}$ can be written as

$$\begin{aligned}\Upsilon_{i,\alpha_1,10} &= \int \frac{\pi_1\bar{\pi}_2h(Y,X,Z,\pi_1)}{a(Y,X_{(-2)},Z,\pi_1)}\left\{S_{\alpha_1}(1,Y,X,Z) - \frac{b_{\alpha_1}(1,Y,X_{(-2)},Z,\pi_1)}{a(Y,X_{(-2)},Z,\pi_1)}\right\}\times I(X_1 = X_{i1}, Z = Z_i)\\ &\quad\times\frac{f(Y|X_{i1},X_2,Z_i)f(X_2|X_{i1},Z_i)}{\mathrm{pr}(R_1 = R_2 = 1|X_{i1},Z_i)}\,\mathrm{d}Y\,\mathrm{d}X_{i2}\\ &= \int \frac{\pi_1\bar{\pi}_2h(Y,X,Z,\pi_1)}{a(Y,X_{(-2)},Z,\pi_1)}\left\{S_{\alpha_1}(1,Y,X,Z) - \frac{b_{\alpha_1}(1,Y,X_{(-2)},Z,\pi_1)}{a(Y,X_{(-2)},Z,\pi_1)}\right\}\times I(X_1 = X_{i1}, Z = Z_i)\\ &\quad\times\frac{f(Y|X_{i1},X_2,Z_i)f(X_2|X_{i1},Z_i,R_1=R_2=1)}{\mathrm{pr}(R_1 = R_2 = 1|X_{i1},X_2,Z_i)}\,\mathrm{d}Y\,\mathrm{d}X_2.\end{aligned}$$

Thus, $\Upsilon_{i,\alpha_1,10}$ can be estimated by replacing $f(X_2|X_{i1}, Z_i, R_1 = R_2 = 1)$ using the empirical density $\widehat{f}(X_2|X_{i1}, Z_i, R_1 = R_2 = 1)$ obtained from the completely observed data. Likewise, for $i \in V$, $\Upsilon_{i,\alpha_1,00}$ can be estimated by

$$\begin{aligned}&\int \frac{\bar{\pi}_1\bar{\pi}_2h(Y,X,Z,\bar{\pi}_1\bar{\pi}_2)}{a(Y,Z,\bar{\pi}_1\bar{\pi}_2)}\left\{S_{\alpha_1}(0,Y,X,Z) - \frac{b_{\alpha_1}(0,Y,Z,\bar{\pi}_1\bar{\pi}_2)}{a(Y,Z,\bar{\pi}_1\bar{\pi}_2)}\right\}\\ &\times\frac{f(Y|X,Z_i)\widehat{f}(X|Z_i,R_1=R_2=1)}{\mathrm{pr}(R_1=R_2=1|X,Z_i)}\,\mathrm{d}Y\,\mathrm{d}X\\ &\times I(Z=Z_i)\frac{f(Y|X,Z_i)\widehat{f}(X|Z_i,R_1=R_2=1)}{\mathrm{pr}(R_1=R_2=1|X,Z_i)}\,\mathrm{d}Y\,\mathrm{d}X.\end{aligned}$$

When Y is continuous, the integration can be evaluated via a quadrature formula. For discrete Y, the integration is replaced by a summation.

4. Extension to More Than Two Covariates

The proposed method with $p = 2$ may be extended to the $p > 2$ scenario by incorporating the missingness mechanism for each of the p missing variables. To estimate conditional distributions, the data can be stratified based on the conditioning variables that include Z and may include a part of X, with the strata (cell) frequencies used for estimating the conditional distributions. However, when the number of strata is large, some cell frequencies could be zero or very small, and consequently the estimates could be unstable. To solve this practical problem with a finite sample size and when the cross classification of the conditioning variables is large, some cells can be merged in a meaningful way, or the following more objective strategy can be adopted: first we seek to identify a set of important variables (they could be a linear combination of the conditioning variables) from the set of the conditioning variables using principal component analysis (PCA). In the second step, based on these identified variables in the first step, we stratify the data. This two-step technique will generally be effective if the dimension of the important variables is less than the dimension of the conditioning variables.

For each partially missing variable X_r $(r = 1, \ldots, p)$, we model $\text{pr}(R_r = 1|Y, X_{(-r)}, Z, \alpha_r)$ parametrically. Suppose that $\mathcal{M}_i$ denotes the set of indices of the missing variables among $X_1, \ldots, X_p$ for the ith observation. Then $\mathcal{M}_i \cup \mathcal{M}_i^c = \{1, \ldots, p\}$. Let the cardinality of $\mathcal{M}_i$ be $|\mathcal{M}_i| = m_i$, and m_i can take values $0, 1, \ldots, p$. For $m_i > 0$, define $X_{i(\mathcal{M}_i)} \equiv \{(X_{ij_1}, \ldots, X_{ij_{m_i}}) : j_1, \ldots, j_{m_i} \in \mathcal{M}_i\}$, and $X_{i(-\mathcal{M}_i)}$ as $X_{i1}, \ldots, X_{ip}$ without $X_{i(\mathcal{M}_i)}$. Then the score function corresponding to the ith observation for which $m_i > 0$ involves an integral with respect to the conditional distribution of $X_{i(\mathcal{M}_i)}$ given $Z_i, X_{i(-\mathcal{M}_i)}$ in the completely observed data $V = \{i : R_{ij} = 1, j = 1, \ldots, p\}$. For empirical estimation of this conditional distribution, we replace $\widehat{f}(X_{i(\mathcal{M}_i)}|Z_i, X_{i(-\mathcal{M}_i)}, R_1 = \cdots = R_p = 1)$ by $\widehat{f}(X_{i(\mathcal{M}_i)}|U_1^{(\mathcal{M}_i)}, \ldots, U_{G_i}^{(\mathcal{M}_i)}, R_1 = \cdots = R_p = 1)$, where $U_1^{(\mathcal{M}_i)}, \ldots, U_{G_i}^{(\mathcal{M}_i)}$ are the first G_i principal components (PCs) of $(Z, X_{(-\mathcal{M}_i)})$ based on the completely observed data. The number of PCs G_i is chosen so that the G_i components capture a high proportion of the total variability of $(Z, X_{(-\mathcal{M}_i)})$. Our numerical experience suggests that at least 80% for the proportion would be sufficient.

Since PCA is used for continuous variables, to apply PCA on a discrete variable, we borrow Kolenikov and Angeles' (2009) idea to first transform the discrete variable into a pseudo-continuous variable as follows. Let X_r be an ordered categorical variable with C_r categories, $0, 1, \ldots, C_r - 1$. Assume that there is a latent variable $L_{X_r} \sim \text{Normal}(0, 1)$ such that X_r takes on 0, c, and $C_r - 1$ when $L_{X_r} \in (-\infty, \kappa_{r1})$, $L_{X_r} \in [\kappa_{rc}, \kappa_{rc+1})$, and $L_{X_r} \in [\kappa_{C_r-1}, \infty)$, respectively, for $c = 1, \ldots, C_r - 2$. The unknown parameters $\kappa_{rc}, c = 1, \ldots, C_r - 1$, are estimated by the maximum likelihood method based on the data where $R_r = 1$. Define the pseudo-continuous variable to be $E(L_{X_r}|X_r = c) = \int_{\kappa_{c-1}}^{\kappa_c} u\phi(u)\,\mathrm{d}u$ when $X_r = c$, where $\phi(\cdot)$ denotes a standard normal density. This technique is one way to reduce the dimension in order to produce a good approximation for our methodology. While it is only an approximate approach and the asymptotic theory does not strictly account for it due to these approximation errors, it is still a practically feasible numerical method to identify important variables. Indeed this dimension reduction technique works well in our simulation scenarios.

The estimating equation for estimating β can now be written compactly

$$S_{\widehat{f}_x,\beta} = \sum_{i=1}^{n} \int S_\beta(Y_i, X_i, Z_i)\widehat{q}_{i,\mathcal{M}_i}(\bar{\pi}_{i(\mathcal{M}_i)}\pi_{i(-\mathcal{M}_i)})\,\mathrm{d}X_{i(\mathcal{M}_i)} = 0,$$

where

$$\widehat{q}_{i,\mathcal{M}_i}(\bar{\pi}_{i(\mathcal{M}_i)}\pi_{i(-\mathcal{M}_i)}) = \left\{ f(Y_i|X_i, Z_i)\bar{\pi}_{i(\mathcal{M}_i)}\pi_{i(-\mathcal{M}_i)} \times \frac{\widehat{f}(X_{i(\mathcal{M}_i)}|U_1^{\mathcal{M}_i}, \ldots, U_{G_i}^{\mathcal{M}_i}, R_{i1} = \cdots = R_{ip} = 1)}{\text{pr}(R_1 = \cdots = R_p = 1|X_i, Z_i)} \right\} \Bigg/ \left\{ \int f(Y_i|X_i, Z_i)\bar{\pi}_{i(\mathcal{M}_i)}\pi_{i(-\mathcal{M}_i)} \times \frac{\widehat{f}(X_{i(\mathcal{M}_i)}|U_1^{\mathcal{M}_i}, \ldots, U_{G_i}^{\mathcal{M}_i}, R_{i1} = \cdots = R_{ip} = 1)}{\text{pr}(R_1 = \cdots = R_p = 1|X_i, Z_i)}\,\mathrm{d}X_{i(\mathcal{M}_i)} \right\}.$$

Note that the integral in the denominator of $\widehat{q}_{i,\mathcal{M}_i}(\bar{\pi}_{i(\mathcal{M}_i)}\pi_{i(-\mathcal{M}_i)})$ is simply a sum over all distinct observed values of $X_{\mathcal{M}_i}$ in the completely observed data. Also, the estimating equation for α_k, $k = 1, \ldots, p$, is

$$S_{\widehat{f}_x,\alpha_k} = \sum_{i=1}^{n} \int S_{\alpha_k}(R_{ik}, Y_i, X_i, Z_i)\widehat{q}_{i,\mathcal{M}_i}(\bar{\pi}_{i(\mathcal{M}_i)}\pi_{i(-\mathcal{M}_i)})\,\mathrm{d}X_{i(\mathcal{M}_i)} = 0,$$

where $\bar{\pi}_{i(\mathcal{M}_i)} = \prod_{j\in\mathcal{M}_i} \text{pr}(R_{ij} = 0|Y_i, X_i, Z_i)$ and $\bar{\pi}_{i(-\mathcal{M}_i)} = \prod_{j\in\mathcal{M}_i^c} \text{pr}(R_{ij} = 1|Y_i, X_i, Z_i)$. Extending the asymptotic arguments given previously, we obtain the influence function representation of $\theta = (\alpha_1^T, \ldots, \alpha_p^T, \beta^T)^{\mathrm{T}}$ with

$$\begin{aligned} S^{\text{adj}}_{\widehat{f}_x,\beta} = &\sum_{i=1}^{n} \int S_\beta(Y_i, X_i, Z_i)q_{i,\mathcal{M}_i}(\bar{\pi}_{i(\mathcal{M}_i)}\pi_{i(-\mathcal{M}_i)})\,\mathrm{d}X_{i(\mathcal{M}_i)} \\ &+ \sum_{i=1}^{n}\sum_{k=1}^{K} \frac{\prod_{j=1}^{p} R_{ij}}{\text{pr}(R_{i1} = \cdots = R_{ip} = 1|Z_i, X_{i(-\mathcal{M}_k)})} \\ &\times E\left[\frac{\bar{\pi}_{i(\mathcal{M}_k)}\bar{\pi}_{i(-\mathcal{M}_k)}h(Y_i, X_i, Z_i, \bar{\pi}_{i(\mathcal{M}_k)}\bar{\pi}_{i(-\mathcal{M}_k)})}{a(Y_i, X_{i(-\mathcal{M}_k)}, Z_i, \bar{\pi}_{i(\mathcal{M}_k)}\bar{\pi}_{i(-\mathcal{M}_k)})} \right. \\ &\times \left\{ S_\beta(Y, X, Z) - \bar{S}_{\beta|\mathcal{M}_k}(Y, Z, X_{(-\mathcal{M}_k)}) \right\} \\ &\times \left. \Big| Z = Z_i, X_{(-\mathcal{M}_k)} = X_{i(-\mathcal{M}_k)} \right], \end{aligned}$$

where $\mathcal{M}_k, k = 1, \ldots, K$, are K distinct patterns of missing data realized in the dataset and $\bar{S}_{\beta|\mathcal{M}_k}(Y, Z, X_{(-\mathcal{M}_k)}) = [\int S_\beta(Y, X, Z)h(X, Y, Z, \bar{\pi}_{\mathcal{M}_k}\pi_{(-\mathcal{M}_k)})f(X_{\mathcal{M}_k}|X_{(-\mathcal{M}_k)}, Z, R_1 = \cdots = R_p = 1)\,\mathrm{d}X_{\mathcal{M}_k}]/a(Y, X_{-(\mathcal{M}_k)}, Z, \bar{\pi}_{\mathcal{M}_k}\pi_{(-\mathcal{M}_k)})$. Similarly,

$$\begin{aligned} S^{\text{adj}}_{\widehat{f}_x,\alpha_l} = &\sum_{i=1}^{n} \int S_{\alpha_l}(R_{il}, Y_i, X_i, Z_i)q_{i,\mathcal{M}_i}(\bar{\pi}_{i(\mathcal{M}_i)}\pi_{i(-\mathcal{M}_i)})\,\mathrm{d}X_{i(\mathcal{M}_i)} \\ &+ \sum_{i=1}^{n}\sum_{k=1}^{K} \frac{\prod_{j=1}^{p} R_{ij}}{\text{pr}(R_{i1} = \cdots = R_{ip} = 1|Z_i, X_{i(-\mathcal{M}_k)})} \\ &\times E\left[\frac{\bar{\pi}_{i(\mathcal{M}_k)}\bar{\pi}_{i(-\mathcal{M}_k)}h(Y_i, X_i, Z_i, \bar{\pi}_{i(\mathcal{M}_k)}\bar{\pi}_{i(-\mathcal{M}_k)})}{a(Y_i, X_{i(-\mathcal{M}_k)}, Z_i, \bar{\pi}_{i(\mathcal{M}_k)}\bar{\pi}_{i(-\mathcal{M}_k)})} \right. \\ &\times \left\{ S_{\alpha_l}(R_{l\mathcal{M}_k}, Y, X, Z) - \bar{S}_{\alpha_l|\mathcal{M}_k}(R_{l\mathcal{M}_k}, Y, Z, X_{(-\mathcal{M}_k)}) \right\} \\ &\times \left. \Big| Z = Z_i, X_{(-\mathcal{M}_k)} = X_{i(-\mathcal{M}_k)} \right], \end{aligned}$$

for $l = 1, \ldots, p$, where $\bar{S}_{\alpha_l|\mathcal{M}_k}(Y, Z, X_{(-\mathcal{M}_k)}) = [\int S_{\alpha_l}(R_{l\mathcal{M}_k}, Y, X, Z)h(X, Y, Z, \bar{\pi}_{\mathcal{M}_k}\pi_{(-\mathcal{M}_k)})f(X_{\mathcal{M}_k}|\ X_{(-\mathcal{M}_k)}, Z, R_1 = \cdots = R_p = 1)\,\mathrm{d}X_{\mathcal{M}_k}]/a(Y, X_{-(\mathcal{M}_k)}, Z, \bar{\pi}_{\mathcal{M}_k}\pi_{(-\mathcal{M}_k)})$.

5. Simulation Studies

5.1. *Simulation Designs*

We considered two scenarios. For scenario I we have two partially missing covariates and one completely observed covariate. We simulated cohort datasets each of size $n = 1000$ by

simulating $Z \sim$ Bernoulli(0.5), $X_1 \sim$ Bernoulli(0.5), and $X_2 \sim$ Bernoulli(0.5). Then the response variable Y was generated from a Bernoulli distribution with $\text{logit}\{\text{pr}(Y = 1|X_1, X_2, Z)\} = -0.5 + 0.2Z + 0.3X_1 - 0.4X_2 - 0.2X_1X_2$. Next we generated the missing values in X_1 and X_2 by randomly simulating binary indicators R_1 and R_2 with $\text{logit}\{\text{pr}(R_1 = 1|Y, Z, X_2)\} = Y + 0.5Z + X_2$ and $\text{logit}\{\text{pr}(R_2 = 1|Y, Z, X_1)\} = Y + 0.5Z + X_1$, respectively. This NI-missingness mechanism yielded approximately 24% missing values in X_1 as well as in X_2. We also considered the MAR mechanism to generate missing values in X_1 and X_2 by simulating R_1 and R_2 from the Bernoulli distribution with success probabilities $\text{logit}\{\text{pr}(R_1 = 1|Y, Z)\} = 0.4 + Y + Z$ and $\text{logit}\{\text{pr}(R_2 = 1|Y, Z)\} = 0.4 + Y + Z$, respectively. In addition, the Supplementary Materials contain some results for NI missing data.

For scenario II, we considered 9 partially missing covariates and one completely observed covariate. We simulated a cohort dataset of size $n = 200$ by simulating $Z \sim$ Bernoulli(0.5), $X_r \sim$ Bernoulli(0.5) for $r = 1, 2, 3, 4$, $X_r \sim$Gamma(2, 2) with $E(X_r) = 1$ for r=5, 6, $X_7 \sim$Normal(0, 1), $X_8 = I(X_1 = 0)$Normal(-0.25, 1) $+ I(X_1 = 1)$Normal(0.25, 1), and $X_9 \sim \text{Normal}(-0.2X_1 + 0.2X_2, 1)$. The response variable Y was taken to be binary with conditional success probability $\text{logit}\{\text{pr}(Y{=}1|Z, X_1, \ldots, X_9)\} = -2 + 0.1Z + 0.2X_1 + 0.3X_2 + 0.4X_3 + 0.5X_4 + 0.5X_5 - 0.5X_6 - 0.4X_7 - 0.3X_8 - 0.2X_9$. The NI-data were generated by simulating the missingness indicators with success probabilities $\text{logit}\{\text{pr}(R_r = 1|Y, Z, X_{(-r)})\} = 0.25 + 0.25Y + 0.25Z + 0.5\sum_{l=1, l\neq r}^{9} X_l$, $r = 1, \ldots, 9$, which resulted in approximately 10–17% missing values for each of X_1 through X_9.

5.2. *Method of Analysis*

The simulated datasets were analyzed by four approaches. First, we analyzed the data without any missing values, where the estimates were obtained by maximizing the full data likelihood $\prod_{i=1}^{n} f(Y_i|X_i, Z_i)$. This approach was referred to as the full data (FD) method. We then analyzed the data using CC, where the subjects containing any missing values in X_k were discarded. The third approach considered here was the extension of the mean-score method (only for scenario I). For this method the underlying assumption was that the data are MAR. Finally, we analyzed the datasets using the proposed semiparametric approach which we referred to as SP. In SP we modeled the missingness mechanism using a linear logistic regression (i.e., the logit of the selection probability is a linear function of the conditioning variables).

5.3. *Results*

Table 1 contains the bias, empirical and estimated standard errors, 95% coverage probabilities based on the Wald-type confidence intervals, and MSE for the five regression parameters involved in the model $f(Y|X_1, X_2, Z)$ for scenario I. The results in Table 1 indicate that when the data are MAR, the mean-score and SP methods performed equally well. Note that both of the approaches have much smaller variances compared with CC. When data are NI-, the proposed approach is much better than the mean-score approach in terms of bias. For the NI-data the mean-score method produces biased results. In both the MAR and NI-scenarios, CC also produced biased results. Moreover, Table 1 shows that the empirical and estimated standard errors are reasonably close for all methods, and the empirical coverage probabilities for SP are close to the nominal level but not always so for the other methods. Additionally, under some NI missing data (presented in the Supplementary Materials), SP shows the least bias among all methods.

For scenario II (Table 2), we only present the results based on CC and the proposed SP approach. We noticed that CC faced some convergence problems (1–2% of the data). The results for CC are based on the datasets where estimates converged and indicate that the proposed approach performs much better than CC in terms of bias and standard errors. While CC shows bias in the parameter estimates, the coverage probabilities for the CC analysis are usually close to the nominal level due to large standard errors. The bias in CC depends on how strongly the missingness mechanism depends on the response variable (results not presented here).

5.4. *Robustness Study for Model Misspecification*

The proposed approach depends on a parametric model assumption of the missing probabilities. Therefore, in principle, a model misspecification can cause bias in the parameter estimates. In order to assess such possible bias, we conducted the following additional simulation with scenario II where NI-missing data were created by simulating the missingness indicators with success probabilities $\text{logit}\{\text{pr}(R_r = 1|Y, Z, X_{(-r)})\} = 0.25 + 0.25Y + 0.25Z + 0.25\sum_{l\neq r}^{9} X_l^2$. The resulting missing percentages for each of the nine variables vary between 9% and 12%. However, in the SP analysis missingness was modeled as $\text{logit}\{\text{pr}(R_r = 1|Y, Z, X_{(-r)})\} = \alpha_{0r} + \alpha_{1r}Y + \alpha_{2r}Z + \sum_{l\neq r}^{9} \alpha_{(2+l),r}X_l$ resulting in a model violation. The results (Table 3) indicate the SP method is hardly affected by this model misspecification. Table 4 contains the results when the actual missingness mechanism was $\text{logit}\{\text{pr}(R_1 = 1|Y, Z)\} = 1 + 0.5Y + 0.5Z$, $\text{logit}\{\text{pr}(R_j = 1|Y, Z, R_{(j-1)})\} = 1.8 + 0.5Y + 0.5Z - 0.8R_{(j-1)}$, $j = 2, \ldots, 9$. This mechanism violates our assumption that the missingness mechanism of each variable is independent conditional on the variables. In this scenario missing percentage for each variable varies between 18% and 21%. In the SP analysis, missingness was modeled as $\text{logit}\{\text{pr}(R_r = 1|Y, Z, X_{(-r)})\} = \alpha_{0r} + \alpha_{1r}Y + \alpha_{2r}Z + \sum_{l\neq r}^{9} \alpha_{(2+l),r}X_l$. The CC analysis had numerical convergence problems for 68% of the datasets out of 500 replications, and here we present the results based only on the converged datasets. There was no convergence issue for SP, which continued to outperform CC.

6. Data Examples

6.1. *Analysis of the Los Angeles Endometrial Cancer Data*

This is a 1:4 matched case-control study of endometrial cancer (Breslow and Day, 1980), where three controls were matched with every case subject based on neighborhood of residence and age in an affluent retirement community in Los Angeles. All subjects in the study were post menopausal women. The dataset contains several risk factors, including presence

Table 1
Results of the simulation study for scenario 1 based on 500 replications

Method	β_0	β_1	β_2	β_3	β_4
FD					
Bias	−0.04	0.05	0.02	0.02	−0.05
EMP.SE	1.52	1.29	1.87	1.90	2.59
EST.SE	1.46	1.32	1.82	1.89	2.65
CP	9.40	9.54	9.50	9.51	9.54
MSE	0.23	0.17	0.35	0.36	0.67
MAR mechanism					
CC					
Bias	5.81	−3.11	0.02	−0.05	0.02
EMP.SE	2.02	1.76	2.34	2.35	3.18
EST.SE	1.95	1.72	2.33	2.38	3.35
CP	1.60	5.61	9.59	9.65	9.61
MSE	3.79	1.26	0.55	0.56	1.01
Mean-score extension					
Bias	−0.01	0.03	0.02	−0.05	0.03
EMP.SE	1.67	1.23	2.27	2.25	3.23
EST.SE	1.73	1.54	2.12	2.19	3.08
CP	9.60	9.82	9.44	9.48	9.38
MSE	0.28	0.17	0.51	0.51	1.04
SP					
Bias	−0.02	0.04	0.01	−0.03	0.02
EMP.SE	1.66	1.29	2.23	2.18	3.18
EST.SE	1.64	1.26	2.31	2.33	3.51
CP	9.48	9.36	9.67	9.69	9.69
MSE	0.27	0.17	0.49	0.48	1.01
NI- mechanism					
CC					
Bias	7.38	−1.62	−1.70	−1.81	0.03
EMP.SE	2.23	1.76	2.65	2.67	3.43
EST.SE	2.22	1.76	2.62	2.66	3.55
CP	0.78	8.46	8.94	9.04	9.59
MSE	5.95	0.57	0.98	1.04	1.17
Mean-score extension					
Bias	1.04	−0.02	−1.03	−1.10	0.35
EMP.SE	1.87	1.30	2.53	2.54	3.45
EST.SE	1.85	1.57	2.24	2.31	3.14
CP	9.24	9.81	8.85	9.11	9.29
MSE	0.46	0.17	0.74	0.76	1.19
SP					
Bias	0.02	0.04	−0.01	−0.09	0.05
EMP.SE	1.77	1.29	2.48	2.47	3.42
EST.SE	1.86	1.25	2.67	2.64	3.66
CP	9.65	9.33	9.70	9.61	9.68
MSE	0.31	0.17	0.61	0.61	1.16

Here FD, CC, SP, EMP.SE, EST.SE, and CP stand for the full data analysis, complete case, the proposed semiparametric method, empirical standard error, estimated standard error, and the 95% coverage probability, respectively. All entries are multiplied by 10.

of gall bladder disease (Z_1), presence of hypertension (Z_2), presence of obesity (X_1), dose of conjugated estrogen used (X_2, denoted as Dose), and duration of conjugated estrogen used (X_3, denoted as Duration). The last three variables had 17%, 3%, and 6% missing values, respectively, while the first two were completely observed. The non-monotone pattern of missing data is given in the top panel of Table 5, and the counts do not show any evidence of dependence between R_1 and R_2, and between R_1 and R_3. Although these counts exhibit strong evidence of dependence between R_2 and R_3, this does not rule out the possibility of independence of these missingness indicators conditional on the covariates. Furthermore, among the subjects with $R_2 \times R_3 = 1$, the missingness of obesity is strongly associated with Duration (p-value= 0.004). Also, among the subjects with $R_1 \times R_2 = 1$, the missingness of Duration is strongly associated with Dose

Table 2
Results of the simulation study based on 500 replications for scenario 2 with NI-mechanism

Method	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
FD											
Bias	−1.77	0.17	0.05	0.36	0.43	0.41	0.46	−0.59	−0.33	−0.27	−0.18
EMP.SE	7.65	4.29	4.34	4.48	4.27	4.42	2.99	3.61	2.24	2.16	2.08
EST.SE	7.16	4.11	4.27	4.15	4.14	4.16	2.77	3.42	2.13	2.09	2.08
CP	9.43	9.53	9.46	9.49	9.62	9.43	9.43	9.57	9.59	9.60	9.51
MSE	6.16	1.89	1.89	2.02	1.85	1.97	0.91	1.38	0.52	0.47	0.44
CC											
Bias	1.58	0.12	−0.46	0.36	0.26	0.49	0.36	−2.12	−1.76	−1.37	−1.11
EMP.SE	15.74	7.43	7.55	7.91	7.55	7.73	4.91	6.90	4.32	4.11	4.14
EST.SE	13.72	6.61	6.91	6.78	6.79	6.87	4.21	5.25	3.64	3.58	3.55
CP	9.22	9.51	9.46	9.26	9.47	9.52	9.46	9.38	9.28	9.32	9.30
MSE	25.00	5.51	5.72	6.27	5.71	6.00	2.42	5.21	2.17	1.87	1.84
SP											
Bias	−0.85	−0.21	−0.23	0.15	−0.09	0.16	0.09	−0.37	−0.11	−0.18	−0.28
EMP.SE	8.19	4.36	4.60	4.54	4.48	4.53	3.12	3.73	2.35	2.13	2.29
EST.SE	7.89	4.23	4.52	4.42	4.39	4.42	3.01	3.67	2.28	2.17	2.24
CP	9.50	9.51	9.52	9.59	9.47	9.47	9.48	9.39	9.55	9.63	9.46
MSE	6.79	1.91	2.12	2.06	2.01	2.05	0.97	0.14	0.55	0.45	0.54

Here FD, CC, SP, EMP.SE, EST.SE, and CP stand for the full data analysis, complete case, the proposed semiparametric method, empirical standard error, estimated standard error, and the 95% coverage probability, respectively. All entries are multiplied by 10.

(p-value= 0.021). These facts motivated us to assume the NI-mechanism and estimate the parameters using the proposed method.

Define $Y = 1$ if a subject has endometrial cancer and $Y = 0$ otherwise. We fit a logistic regression model $\text{logit}\{\text{pr}(Y = 1|Z_1, Z_2, X_1, X_2, X_3)\} = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 X_1 + \beta_4 X_2 + \beta_5 X_3$ as if the data were unmatched and cohort. We estimated the model parameters using CC and SP. For the missingness model in SP we considered a logistic regression model, that is, $\text{logit}\{\text{pr}(R_r = 1|X_{(-r)}, Z, Y)\} = \alpha_{r0} + \alpha_{r1} Y + \sum_{j=1}^{2} \alpha_{r(1+j)} Z_j + \sum_{j=1, j\neq r}^{3} \alpha_{r(3+j)} X_j$. The estimates, standard errors, and p values for β from both methods are presented

Table 3
Results of the simulation study based on 500 replications for scenario 2 with NI-mechanism with $logit\{pr(R_r = 1|Y, Z, X_{(-r)})\} = 0.25 + 0.25Y + 0.25Z + 0.25 \sum_{l\neq r}^{9} X_l^2$

Method	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
FD											
Bias	−1.90	0.10	0.37	0.15	0.70	0.38	0.45	−0.64	−0.38	−0.28	−0.36
EMP.SE	7.86	4.53	4.60	4.39	3.99	4.36	3.06	3.67	2.23	2.03	2.29
EST.SE	7.19	4.14	4.28	4.16	4.16	4.18	2.78	3.43	2.13	2.09	2.09
CP	9.40	9.44	9.42	9.36	9.70	9.54	9.40	9.40	9.48	9.68	9.36
MSE	6.53	2.05	2.13	1.93	1.64	1.91	0.95	1.39	0.50	0.42	0.54
CC											
Bias	0.37	−0.24	0.09	0.17	1.01	0.50	0.96	−2.83	−2.11	−0.73	−0.69
EMP.SE	15.25	8.32	13.96	8.66	7.81	9.62	8.87	7.39	7.31	3.88	5.02
EST.SE	14.36	7.35	7.89	7.45	7.42	7.57	4.61	5.58	3.74	3.36	3.37
CP	9.67	9.30	9.44	9.65	9.56	9.40	9.42	9.24	9.59	9.32	9.22
MSE	23.23	6.91	19.47	7.50	6.18	9.25	7.95	6.26	5.77	1.55	2.56
SP											
Bias	−1.55	0.01	−0.03	−0.08	0.44	0.14	0.05	−0.18	0.03	0.14	0.32
EMP.SE	8.11	4.48	4.63	4.54	4.19	4.46	3.03	3.79	2.29	2.18	2.23
EST.SE	7.57	4.19	4.42	4.31	4.31	4.33	2.88	3.44	2.18	2.14	2.13
CP	9.44	9.46	9.44	9.42	9.64	9.50	9.42	9.32	9.44	9.54	9.44
MSE	6.80	2.00	2.14	2.06	1.77	1.98	0.91	1.43	0.52	0.47	0.49

For the SP method the missingness mechanism is incorrectly modeled. Here FD, CC, SP, EMP.SE, EST.SE, and CP stand for the full data analysis, complete case, the proposed semiparametric method, empirical standard error, estimated standard error, and the 95% coverage probability, respectively. All entries are multiplied by 10.

Table 4

Results of the simulation study based on 500 replications for scenario 2 where the missingness mechanism is $logit\{pr(R_1 = 1|Y, Z)\} = 1 + 0.5Y + 0.5Z$, $logit\{pr(R_j = 1|Y, Z, R_{(j-1)})\} = 1.8 + 0.5Y + 0.5Z - 0.8R_{(j-1)}$, $j = 2, \ldots, 9$

Method	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
FD											
Bias	−1.93	−0.15	0.34	0.58	0.49	0.36	0.49	−0.75	−0.25	−0.31	−0.37
EMP.SE	7.86	4.65	4.68	4.53	4.03	4.42	2.96	3.62	2.18	2.15	2.28
EST.SE	7.19	4.13	4.28	4.17	4.16	4.18	2.76	3.44	2.12	2.10	2.09
CP	9.52	9.30	9.30	9.36	9.64	9.56	9.40	9.60	9.52	9.56	9.40
MSE	6.55	2.16	2.20	2.08	1.64	1.95	0.90	1.36	0.48	0.47	0.53
CC											
Bias	9.81	−5.44	−0.95	1.19	1.01	2.11	2.99	−4.88	−1.52	−1.92	−1.23
EMP.SE	24.92	15.23	15.17	15.82	13.60	15.77	11.43	12.28	8.49	8.46	10.78
EST.SE	26.84	16.46	15.47	15.07	15.04	15.37	11.38	13.13	7.82	8.05	8.42
CP	9.62	9.81	9.91	9.68	9.82	9.82	9.91	9.88	9.75	9.82	9.69
MSE	71.38	26.03	22.93	25.03	18.48	25.17	13.86	17.38	7.39	7.50	11.71
SP											
Bias	−2.03	0.07	−0.10	0.53	0.28	−0.02	−0.05	−0.20	0.27	0.15	−0.05
EMP.SE	8.13	4.64	4.61	4.64	4.38	4.65	3.17	4.00	2.28	2.21	2.43
EST.SE	7.58	4.20	4.52	4.43	4.43	4.42	2.97	3.66	2.26	2.22	2.23
CP	9.52	9.38	9.54	9.46	9.66	9.54	9.40	9.30	9.40	9.56	9.32
MSE	7.01	2.15	2.13	2.18	1.92	2.16	1.00	1.60	0.52	0.49	0.59

For the SP method the missingness mechanism is incorrectly modeled. Here FD, CC, SP, EMP.SE, EST.SE, and CP stand for the full data analysis, complete case, the proposed semiparametric method, empirical standard error, estimated standard error, and the 95% coverage probability, respectively. All entries are multiplied by 10.

in Table 5. SP has lower standard errors for all the coefficient estimates. Except for X_1 and X_3, the parameter estimates for both the methods are somewhat similar. Further investigation revealed that the missingness mechanism of X_1 strongly depends on X_3 (p-value = 0.005) and the missingness mechanism of X_3 strongly depends on X_1 (p-value = 0.034). Based on the results of SP, the presence of gall bladder disease and duration of conjugated estrogen used are statistically significant (at the 5% level) as risk factors for endometrial cancer.

Table 5

The top panel shows the pattern of missing data in the Los Angeles Endometrial Cancer data with 1 and 0 representing the observed and missing variable, respectively, and the bottom panel contains the results of the data analysis

Pattern	Obesity	Dose of conjugated estrogen	Duration of conjugated estrogen used	Count	
1	0	0	0	1	
2	0	1	0	1	
3	1	0	0	4	
4	1	1	0	11	
5	0	0	1	1	
6	0	1	1	48	
7	1	0	1	2	
8	1	1	1	247	
Method	Gall bladder	Hypertension	Obesity	Dose	log (duration)
CC					
Estimate	1.143	0.083	0.743	0.269	0.224
SE	0.437	0.352	0.391	0.120	0.103
p-Value	0.009	0.815	0.057	0.025	0.029
SP					
Estimate	1.164	0.098	0.459	0.202	0.298
SE	0.420	0.312	0.347	0.105	0.091
p-Value	0.005	0.753	0.185	0.055	0.001

Here CC and SP stand for the complete case and the proposed semiparametric method, respectively.

Table 6
Results of the hip fracture data analysis

Method	CC			SP		
	Estimate	SE	p-Value	Estimate	SE	p-Value
EtOH	1.391	0.391	<0.001	1.367	0.336	<0.001
Smoke	0.929	0.400	0.020	0.913	0.342	0.007
Dementia	2.509	0.724	<0.001	2.472	0.724	<0.001
Antiseiz	3.310	1.064	0.002	3.293	0.897	<0.001
LevoT4	2.010	1.015	0.047	1.970	0.795	0.013
AntiChol	−1.918	0.768	0.012	−1.914	0.757	0.011
Albumin	−0.911	0.353	0.009	−0.964	0.293	0.001
BMI	−0.104	0.038	0.006	−0.093	0.030	0.002
log(HGB)	−2.596	1.202	0.031	−2.618	0.943	0.005

Here CC and SP stand for the complete case and the proposed semiparametric method, respectively.

6.2. *Analysis of the Hip Fracture Data*

The second data example is a 1:1 age and race matched case-control study of possible risk factors of hip fracture among male veterans with 218 strata (Chen, 2004). The study was conducted at the University of Illinois at Chicago College of Medicine (Barengolts et al., 2001). Many potential risk factors (25 altogether) in addition to age and race were recorded. Preliminary exploratory analysis suggested that nine of these risk factors are potentially important. Among them EtOH(X_1), Smoke (X_2), Dementia (X_3), Antiseizure (X_4), LevoT4 (X_5), and anti-cholesterol (X_6) are dichotomous variables, whereas the remaining three Albumin (X_7), BMI (X_8), and log(HGB) (X_9) are continuous in nature. All of these predictors contain some missing values, with the percentage of missing values being approximately 10%, 12%, 3%, 5%, 9%, 10%, 24%, 10%, and 12%, respectively. Although there could be at most $2^9 - 1 = 511$ missing patterns, the observed number of patterns (non-monotone) is 37, and these patterns are presented in Table 2 of Chen (2004). The presence ($Y = 1$) or absence ($Y = 0$) of hip fracture was considered as the binary outcome. We aimed to find the estimate of the log-odds ratio parameters by fitting a logistic regression model to the data: $\text{logit}\{\text{pr}(Y = 1|X)\} = \beta_0 + \sum_{r=1}^{9} X_r\beta_r$. Note here that $\beta_1, \ldots, \beta_9$ are the main parameters of interest. Following Chen (2004), in the model we ignored age and race as these variable were used as matching variables.

First we carried out the CC analysis and then we analyzed the data using SP. In the latter, we assumed that the missingness mechanisms follow a logistic model: $\text{logit}\{\text{pr}(R_r = 1|X_{(-r)}, Y)\} = \alpha_{r0} + \alpha_{r1}Y + \sum_{j=1, j\neq r}^{9} \alpha_{r(j+1)}X_j$ for $r = 1, \ldots, 9$. We applied the dimension reduction technique for calculating the conditional distributions based on the completely observed data. In Table 6 we present the estimates, standard errors, and p values for β-parameters. The estimates differ somewhat between the methods. However, the standard errors of SP are substantially smaller than that of CC. The results indicate that all the covariates have statistically significant associations (at the 5% level) with the occurrence of hip fracture. Our results are consistent with the results of Chen (2004) who analyzed this data under the MAR assumption.

7. Discussion

This article deals with non-monotone patterns of missing covariate data in a parametric regression model. Here we address two important issues simultaneously: extending the missingness mechanism beyond MAR and handling possible association among the partially missing covariates. The method relies on a parametric model for the missingness mechanism, and possible associations among the covariates are handled empirically without using any parametric structure. The NI-mechanism may be used when the MAR assumption is likely to be violated (Potthof et al., 2006) or in the scenario when the NI missing assumption is plausible. If we assume that the missingness mechanism is completely non-ignorable, then we will require a parametric model for the partially missing covariate. This parametric modeling of the missing covariate distribution is avoided by imposing the NI-mechanism assumption.

The simulation results indicate that the proposed approach works remarkably well in terms of bias and standard error. Furthermore, it appears robust when some of the underlying assumptions about the missingness mechanism are violated. We have also derived the asymptotic normality of the proposed estimator and an asymptotically consistent formula to compute the standard errors.

Our proposed method is an estimated-score approach to estimate the parameters of the model. Alternatively, a maximum likelihood estimator (MLE) or a nonparametric maximum likelihood estimator (NPMLE) is also possible and likely to be more efficient. However, in both cases of MLE and NPMLE the computation could be challenging for a large dimension of the parameter vector as all parameters (β, $\alpha_1, \ldots, \alpha_p$, and any additional parameters) need to be estimated simultaneously in each step of the EM algorithm. In contrast, in each iteration of our EM algorithm, β, $\alpha_1, \ldots, \alpha_p$ are estimated separately. Thus, a further theoretical and numerical investigation of the MLE-type estimator would be interesting. Finally, in principle the proposed idea can be generalized to handle both missing covariates and the missing response variable, especially in the context of longitudinal data analyses.

8. Supplementary Materials

Some technical details for the identifiability result in Section 2.3, the proof of Theorem 1 along with the regularity conditions, one simulation table, and the code for the simulation studies are available with this paper at the *Biometrics* website on Wiley Online Library.

Acknowledgements

The authors thank the Editor, an Associate Editor, and two referees for their constructive and helpful comments and suggestions which have led to a much improved version of the manuscript, and Hua Yun Chen for kindly providing the hip fracture dataset. This research was partially supported by NSF grant SES 0961618, NIH grant R03CA176760, and Award Number KUS-CI-016-04 made by King Abdullah University of Science and Technology (KAUST). Part of the work was carried out while Wang was visiting Australian National University supported by the Mathematical Sciences Research Visitor Program.

Appendix: Expressions for the adjustment terms

$$\Upsilon_{i,\alpha_1,10}=\frac{R_{i1}R_{i2}}{\mathrm{pr}(R_1=R_2=1|X_{i1},Z_i)}E\left[\frac{\pi_1\bar{\pi}_2h(Y,X,Z,\pi_1)}{a(Y,X_{(-2)},Z,\pi_1)}\times\left\{S_{\alpha_1}(1,Y,X,Z)-\frac{b_{\alpha_1}(1,Y,X_{(-2)},Z,\pi_1)}{a(Y,X_{(-2)},Z,\pi_1)}\right\}\times|X_1=X_{i1},Z=Z_i\right],$$

$$\Upsilon_{i,\alpha_1,00}=\frac{R_{i1}R_{i2}}{\mathrm{pr}(R_1=R_2=1|Z_i)}E\left[\frac{\bar{\pi}_1\bar{\pi}_2h(Y,X,Z,\bar{\pi}_1\bar{\pi}_2)}{a(Y,Z,\bar{\pi}_1\bar{\pi}_2)}\times\left\{S_{\alpha_1}(0,Y,X,Z)-\frac{b_{\alpha_1}(0,Y,Z,\bar{\pi}_1\bar{\pi}_2)}{a(Y,Z,\bar{\pi}_1\bar{\pi}_2)}\right\}\Bigg|Z=Z_i\right],$$

$$\Upsilon_{i,\alpha_2,01}=\frac{R_{i1}R_{i2}}{\mathrm{pr}(R_1=R_2=1|X_{i2},Z_i)}E\left[\frac{\bar{\pi}_1\pi_2h(Y,X,Z,\pi_2)}{a(Y,X_{(-1)},Z,\pi_2)}\times\left\{S_{\alpha_2}(1,Y,X,Z)-\frac{b_{\alpha_2}(1,Y,X_{(-1)},Z,\pi_2)}{a(Y,X_{(-1)},Z,\pi_2)}\right\}\times\Big|X_2=X_{i2},Z=Z_i\right],$$

$$\Upsilon_{i,\alpha_2,00}=\frac{R_{i1}R_{i2}}{\mathrm{pr}(R_1=R_2=1|Z_i)}E\left[\frac{\bar{\pi}_1\bar{\pi}_2h(Y,X,Z,\bar{\pi}_1\bar{\pi}_2)}{a(Y,Z,\bar{\pi}_1\bar{\pi}_2)}\times\left\{S_{\alpha_2}(0,Y,X,Z)-\frac{b_{\alpha_2}(0,Y,Z,\bar{\pi}_1\bar{\pi}_2)}{a(Y,Z,\bar{\pi}_1\bar{\pi}_2)}\right\}\Bigg|Z=Z_i\right],$$

$$\Upsilon_{i,\beta,10}=\frac{R_{i1}R_{i2}}{\mathrm{pr}(R_1=R_2=1|X_{i1},Z_i)}E\left[\frac{\pi_1\bar{\pi}_2h(Y,X,Z,\pi_1)}{a(Y,X_{(-2)},Z;\pi_1)}\times\left\{S_{\beta}(Y,X,Z)-\frac{a_{\beta}(Y,X_1,Z;\pi_1)}{a(Y,X_{(-2)},Z;\pi_1)}\right\}\Big|X_1=X_{i1},Z=Z_i\right],$$

$$\Upsilon_{i,\beta,01}=\frac{R_{i1}R_{i2}}{\mathrm{pr}(R_1=R_2=1|X_{i2},Z_i)}E\left[\frac{\bar{\pi}_1\pi_2h(Y,X,Z,\pi_2)}{a(Y,X_{(-1)},Z;\pi_2)}\times\left\{S_{\beta}(Y,X,Z)-\frac{a_{\beta}(Y,X_2,Z;\pi_2)}{a(Y,X_{(-1)},Z;\pi_2)}\right\}\Big|X_2=X_{i2},Z=Z_i\right],$$

$$\Upsilon_{i,\beta,00}=\frac{R_{i1}R_{i2}}{\mathrm{pr}(R_1=R_2=1|Z_i)}E\left[\frac{\bar{\pi}_1\bar{\pi}_2h(Y,X,Z,\bar{\pi}_1\bar{\pi}_2)}{a(Y,Z;\bar{\pi}_1\bar{\pi}_2)}\times\left\{S_{\beta}(Y,X,Z)-\frac{a_{\beta}(Y,Z;\bar{\pi}_1\bar{\pi}_2)}{a(Y,Z;\bar{\pi}_1\bar{\pi}_2)}\right\}\Bigg|Z=Z_i\right],$$

where $b_{\alpha_1}(R_1,Y,X_{(-r)},Z,\omega)=\int S_{\alpha_1}(R_1,Y,X,Z)h(Y,X,Z,\omega)f(X_r|X_{(-r)},Z,R_1=R_2=1)\,\mathrm{d}X_r$, $b_{\alpha_2}(R_2,Y,X_{(-r)},Z,\omega)=\int S_{\alpha_2}(R_2,Y,X,Z)h(Y,X,Z,\omega)f(X_r|X_{(-r)},Z,R_1=R_2=1)\,\mathrm{d}X_r$, $a(Y,X_{(-r)},Z;\omega)=\int h(Y,X,Z,\omega)f(X_r|X_{(-r)},Z,R_1=R_2=1)\,\mathrm{d}X_r$, $a(Y,Z;\omega)=\int h(Y,X,Z,\omega)f(X|Z,R_1=R_2=1)\,\mathrm{d}X$, $a_{\beta}(Y,X_{(-r)},Z;\omega)=\int S_{\beta}(Y,X,Z)h(Y,X,Z,\omega)f(X_r|X_{(-r)},Z,R_1=R_2=1)\,\mathrm{d}X_r$, $a_{\beta}(Y,Z;\omega)=\int S_{\beta}(Y,X,Z)h(Y,X,Z,\omega)f(X|Z,R_1=R_2=1)\,\mathrm{d}X$.

References

Barengolts, E., Karanouh, D., Kolodny, L., and Kukreja, S. (2001). Risk factors for hip fractures in predominantly AfricanAmerican veteran male population. *Journal of Bone and Mineral Research* **16**, S170.

Breslow, N. E. and Day, N. E. (1980). *Statistical Methods in Cancer Research*, Vol. 1. Lyon, France: International Agency for Research on Cancer.

Chatterjee, N., Chen, Y.-H., and Breslow, N. E. (2003). A pseudoscore estimator for regression problems with two-phase sampling *Journal of the American Statistical Association* **98**, 158–168.

Chen, H.-Y. (2004). Nonparametric and semiparametric models for missing covariates in parametric regression. *Journal of the American Statistical Association* **99**, 1176–1189

Ibrahim, J. G, Chen, M. H, and Lipstiz, S. R. (1999a). Monte Carlo EM for missing covariates in parametric regression models. *Biometrics* **55**, 591–596.

Ibrahim, J. G., Lipsitz, S. R., and Chen, M.-H. (1999b). Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *Journal of the Royal Statistical Society, Series B* **61**, 173–190.

Kolenikov, S. and Angeles, G. (2009). Socioeconomic status measurement with discrete proxy variables: Is principal component analysis a reliable answer? *Review of Income and Wealth* **55**, 128–165.

Lei, S.-Y. C. and Wang, S. (2001). Diagnostic tests for bias of estimating equations in weighted regression with missing covariates. *Canadian Journal of Statistics* **29**, 239–250.

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

Little, R. J. A. (1995). Modeling the drop-out mechanism in longitudinal studies. *Journal of the American Statistical Association* **90**, 1112–1121.

Potthoff, R. A., Tudor, G. E., Pieper, K. S., and Hasselblad, V. (2006). Can one assess whether missing data are missing at

random in medical studies. *Statistical Methods in Medical Research* **15**, 213–234.

Reilly, M. and Pepe, M. S. (1995). A mean-score method for missing and auxiliary covariate data in regression models. *Biometrika* **82**, 299–314.

Robins, J. M. (1997). Non-response models for the analysis of non-monotone non-ignorable missing data. *Statistics in Medicine* **16**, 21–37.

Robins, J. M. and Gill, R. (1997). Non-response models for the analysis of non-monotone ignorable missing data. *Statistics in Medicine* **16**, 39–56.

Robins, J. M., Rotnitzky, A., and Zhao L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846–866.

Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica* **39**, 577–591.

Rubin, D. B. (1976). Inference and missing data (with discussion). *Biometrika* **63**, 581–592.

Received June 2013. *Revised January* 2014.
Accepted February 2014.