A self-censoring model for multivariate nonignorable nonmonotone missing data

Yilin Li | Wang Miao | Ilya Shpitser | Eric J. Tchetgen Tchetgen

1Department of Probability and Statistics, Peking University, Beijing, China
2Department of Computer Science, Johns Hopkins University, Baltimore, USA
3Department of Statistics, The Wharton School of the University of Pennsylvania, Philadelphia, Pennsylvania, USA

Abstract
We introduce an itemwise modeling approach called “self-censoring” for multivariate nonignorable nonmonotone missing data, where the missingness process of each outcome can be affected by its own value and associated with missingness indicators of other outcomes, while conditionally independent of the other outcomes. The self-censoring model complements previous graphical approaches for the analysis of multivariate nonignorable missing data. It is identified under a completeness condition stating that any variability in one outcome can be captured by variability in the other outcomes among complete cases. For estimation, we propose a suite of semiparametric estimators including doubly robust estimators that deliver valid inferences under partial misspecification of the full-data distribution. We also provide a novel and flexible global sensitivity analysis procedure anchored at the self-censoring. We evaluate the performance of the proposed methods with simulations and apply them to analyze a study about the effect of highly active antiretroviral therapy on preterm delivery of HIV-positive mothers.

KEYWORDS
doubly robust estimation, identification, missing not at random, nonmonotone missingness

1 INTRODUCTION
Missing data arise ubiquitously in empirical studies. The missing data mechanism is called missingness at random or ignorable if it does not depend on the missing variables conditional on fully-observed ones. Otherwise, it is called missingness not at random (MNAR) or nonignorable (Rubin, 1976). Nonignorable missingness introduces fundamental challenges to identification, where identification means that the full-data distribution or a statistical functional of interest is uniquely determined from the observed-data distribution. For a single missing outcome, Wang et al. (2014) and Miao et al. (2016) illustrated that identification is not ensured even for fully parametric models, and fully-observed auxiliary variables such as shadow variables (D’Haultfoeuille, 2010; Miao & Tchetgen Tchetgen, 2016) or instrumental variables (Sun et al., 2018) have been used to achieve identification under nonignorable missingness. However, for multivariate missing data, the missingness of each variable may depend on different partially-observed variables and the missingness of other variables, which leads to difficulty for specifying provably identifiable models. For multivariate missing data, the missingness is called monotone if the variables can be ordered such that once a variable is unobserved, all variables later in the order are also missing. This occurs frequently in longitudinal studies where missingness is due to dropout. Otherwise, arbitrary patterns of missingness can arise, which is called nonmonotone missingness.
Previous researchers have studied a variety of models for nonignorable nonmonotone missingness, including the group permutation model (Robins, 1997) which states that missingness can depend on “previous” and “observed future” but not on the present outcome, and the block-conditional missing at the random model (Zhou et al., 2010) which characterizes the potentially nonignorable missingness for blocks of variables in a sequential way. Tchetgen Tchetgen et al. (2018) introduced discrete choice models which connect the multinomial missingness mechanisms to some underlying utility functions. Linero (2017) proposed a nearest identified pattern restriction for identification under Bayesian inference framework. Du et al. (2022) developed a Bayesian latent variable approach by separately modeling the missingness of each outcome with a probit model. Directed acyclic graphs (DAGs) are a useful tool for describing multivariate missingness and studying identification. Fay (1986) and Ma et al. (2003) studied identification of DAGs with categorical outcomes. Mohan and Pearl (2021) provided a comprehensive review on missing data methods using graphical models, established conditions for identification and proposed testable implications for DAGs. Nabi et al. (2020) established necessary and sufficient identifying conditions for the full-data distribution when the restrictions on the full-data distribution are implied by its factorization with respect to a DAG. Notably, the no self-censoring (NSC) model was recently proposed for the analysis of multivariate nonignorable nonmonotone missing data (Sadinle & Reiter, 2017; Shpitser, 2016; Malinsky et al., 2022), which assumes that the missingness process of each outcome does not depend on its own value after conditioning on the other outcomes and missingness indicators. In general, it depends on the amount of information from data collecting process and observed data to decide which missingness assumption is most plausible.

It is commonly encountered in practice that the missingness of an outcome is affected by its own value. This is plausible in situations where the failure of data collection, data loss, or units’ unwillingness to respond is affected by data values. Examples include religious beliefs and sexual preferences in epidemiological studies, smokers not reporting their smoking behavior in insurance applications, and voters not disclosing their political preferences in election surveys. Although this notion of self-censoring for a univariate outcome has been studied extensively (e.g., D’Haultfœuille, 2010; Wang et al., 2014; Miao et al., 2016; Sun et al., 2018), the multivariate self-censoring mechanism is rarely studied with only few exceptions. Brown (1990) considered a self-censoring mechanism for multivariate normal outcomes with a fully-observed outcome. The self-censoring mechanism enjoys the ease for interpretation to applied researchers and admits a sensitivity analysis based on expert substantive knowledge.

In this paper, we make several methodological contributions. First, we describe a self-censoring model for multivariate nonignorable nonmonotone missing data. The self-censoring model reveals that the missingness process of each outcome can depend on its underlying value, the missingness status of other outcomes, and some possibly fully-observed covariates, but not on other partially-observed outcomes. Then, we establish the identification condition for the full-data distribution without imposing any parametric restriction, but invoke a completeness condition that is commonly used in nonparametric identification problems, which accommodates both discrete and continuous variables. A generalized odds ratio parameterization is adopted to facilitate estimation by factorizing the full-data distribution into a baseline propensity score model, a baseline outcome model, and an odds ratio function. Under this parameterization, we develop a suite of semiparametric estimation methods including an inverse probability weighted (IPW) estimator and an outcome regression/imputation-based estimator, which require correct specification of the respective baseline propensity score or the baseline outcome model, together with the odds ratio function. To promote robustness against model misspecification, we propose a doubly robust estimator, which is consistent as long as the odds ratio model and at least one of the baseline models are correct. We also provide a novel procedure for sensitivity analysis anchored at the self-censoring assumption. We evaluate performance of the proposed estimation methods as well as the sensitivity of our methods against misspecification with simulations and apply them to a real-data problem about the effect of continuing highly active antiretroviral therapy (HAART) on preterm delivery of mothers with HIV-positive and maternal hypertension in Botswana. Our analysis approach indicates that continuing HAART increases the risk of preterm delivery.

2 | THE SELF-CENSORING MODEL

2.1 | Model assumptions

Let \( \mathbf{Y} = (Y_1, \ldots, Y_p) \) denote a vector of outcomes subject to missingness, \( \mathbf{R} = (R_1, \ldots, R_p) \) the vector of missingness indicators or patterns with \( R_i = 1 \) if \( Y_i \) is observed and \( R_i = 0 \) otherwise, and \( \mathbf{X} \) a vector of \( d \) fully-observed covariates. We use capital letters for random variables and lowercase letters for realization of the corresponding variables. The observed data are comprised of independent and identically distributed realizations of \( (\mathbf{Y}(\mathbf{R}), \mathbf{R}, \mathbf{X}) \), where \( \mathbf{Y}(\mathbf{R}) = \{ Y_i : R_i = 1 \} \) denotes the observed outcomes of \( \mathbf{Y} \) for pattern \( \mathbf{R} \). Let \( \mathcal{R} = \{ \mathbf{r} \in \{0,1\}^p : p(\mathbf{r}) > 0 \} \) be the set of all missingness patterns. We use \( \mathbf{R} = \mathbf{r} \) as shorthand for
\((R_1, \ldots, R_p) = (r_1, \ldots, r_p), R = 1\) for \(R = (1, \ldots, 1)\), and likewise \(R = 0\) for \(R = (0, \ldots, 0)\). For a vector \(V\), let \(V_i\) denote the subvector after removing \(V_i\). The following assumption characterizes a self-censoring mechanism for multivariate missing data.

**Assumption 1.** (i) For \(i = 1, \ldots, p, R_i \perp Y_i | (X, Y_i, R_{\cdot \cdot})\); (ii) For any \(r \in R, p(r|x, y) > c\) almost surely for some constant \(c > 0\), and if \(r_i \leq r'_i\) for all \(i\) then \(r' \in R\).

Assumption 1 describes a nonignorable missingness mechanism for multivariate outcomes. Condition (i) reveals a mechanism that the missingness of each outcome is not affected by the value of other outcomes conditional on the missingness indicators of other outcomes, and fully-observed covariates. However, the missingness of each outcome can be affected by the value of itself, thus referred to as self-censoring. This condition does not exclude marginal dependencies between \(R_i\) and \(Y_i\). A special case is that \(R_i \perp (Y_i, R_{\cdot \cdot}) | (X, Y_i)\) for \(i = 1, \ldots, p\), which characterizes a strict self-censoring model where the missingness of each outcome can only depend on the outcome’s value and fully-observed covariates, also known as the generalized censoring mechanism by Brown (1990). There exist situations where the self-censoring mechanism is deemed to be more plausible, particularly in surveys with certain sensitive questions (e.g., HIV status, income, or drug use) so that social stigma makes the nonresponse of outcomes directly dependent on their underlying values. Condition (i) imposes restrictions on the observed-data distribution. Therefore, part of the conditional independence can in principle be refuted with the observed data discussed in Web Appendix A. The first part of condition (ii) is a common positivity condition stating that the propensity for any missingness pattern is bounded away from zero. It is necessary for identification of the full-data distribution and consistent estimation of its functionals. If a missingness pattern \(r\) exists, the second part of (ii) requires the existence of missingness patterns \(r'\) with more observed outcomes; this condition rules out certain missingness mechanisms such as monotone missingness. For example, if \(r = (1, 0, 0) \in R\), then condition (ii) requires that both \(r' = (1, 0, 1)\) and \((1, 1, 0) \in R\), in which case monotone missingness does not hold.

A notable model for multivariate nonignorable missing outcomes is the NSC model previously studied by Sadinle and Reiter (2017), Shpitser (2016), and Malinsky et al. (2022), which assumes that \(R_i \perp Y_i | (X, Y_i, R_{\cdot \cdot})\) for \(i = 1, \ldots, p\), that is, the missingness of each outcome is not influenced by its own value but by the remaining partially-observed outcomes. The model in Assumption 1 complements by considering the self-censoring mechanism. It has been noted that if self-censoring exists the full-data distribution is unlikely to be identified without additional restrictions (e.g., Nabi et al., 2020; Mohan & Pearl, 2021); however, in the next section we show that the full-data distribution is identifiable for the self-censoring model characterized in Assumption 1 under a widely-used completeness condition. The self-censoring model can be viewed as a generalization of the shadow variable approach (Miao & Tchetgen Tchetgen, 2016; D’Haultfoeuille, 2010) by admitting missing outcomes as shadow variables and identifying multivariate missing outcomes in a unifying way. Under the self-censoring model, \(Y_i\) serves as a vector of shadow variables for each \(Y_i\), which is independent of \(R_i\) conditional on \((X, Y_i, R_{\cdot \cdot}) = 1\). Previous shadow variable approaches typically require the shadow variable to be fully-observed; however, here \(Y_i\) is prone to nonignorable missingness, which further complicates identification and inference in ways not addressed by existing shadow variable literature. We extend the self-censoring model to blockwise in Web Appendix C, which allows arbitrary missingness mechanism within each block and a self-censoring mechanism across blocks.

Although we do not emphasize graphical concepts, graphical models can provide insight and intuition for identification and analysis for multivariate missing data. The self-censoring model can also be encoded in a chain graph \(G = (V, E)\), where the set of vertices \(V\) is comprised of \((X, Y, R)\) and the set of edges \(E\) contains directed edges from each \(Y_i\) to \(R_i\), directed edges from \(X\) to \((Y, R)\), directed edges among \(Y\), and undirected edges among \(R\). The directed edges among \(Y\) may be reversed without altering the self-censoring model. See Lauritzen (1996) for details of chain graphs. Figure 1a presents an example chain graph for a self-censoring model with three outcomes. We characterize the conditional independence by a chain graph instead of the DAG, because the edges among \(R\) are not allowed to be directed under the self-censoring assumption. For strict self-censoring model, this chain graph reduces to the DAG in Figure 1b, where the edges among \(R\) disappear.

### 2.2 Identification

In this section, we suppress observed covariates for notation simplicity, and all results can be viewed as conditioning on them. The joint distribution \(p(y, r)\) is identified if and only if the missingness mechanism \(p(r|y)\) is identified by \(p(y, r) = p(y, r = 1)p(r|y)/p(r = 1|y)\). To identify \(p(r|y)\), we first split it into several parts by the odds ratio factorization (Chen, 2010; Malinsky et al., 2022) which holds under any missingness mechanism,

\[
p(r|y) = \frac{\prod_{i=1}^{p} p(r_i|y, r_{\cdot i} = 1) \prod_{j=1}^{p} \eta_i(r_{i}, r_{\cdot i}, y)}{\sum_{r' \in R} \prod_{i=1}^{p} p(r'_i|y, r_{\cdot i} = 1) \prod_{j=2}^{p} \eta_i(r'_i, r'_{\cdot i}, y)},
\]  

(1)
where \( \eta_i \) is the sequential odds ratio function between the \( i \)th and the preceding missingness indicators. Letting \( r_{<i} = (r_1, \ldots, r_{i-1}) \) and \( r_{\ge i} = (r_{i+1}, \ldots, r_p) \), we have

\[
\eta_i(r_i, r_{<i}, y) = \frac{p(r_i|y, r_{<i}, r_{\ge i} = 1)p(r_1 = 1|y, r_\le i = 1)}{p(r_i = 1|y, r_{<i}, r_{\ge i} = 1)p(r_1|y, r_\le i = 1)},
\]

quantifying the association between the missingness indicators; for the strict self-censoring model, \( \eta_i(r_i, r_{<i}, y) = 1 \). Although we express the missingness mechanism by the sequential odds ratio factorization, the term \( \prod_{i=2}^p \eta_i(r_i, r_{<i}, y) \) can be represented in terms of pairwise odds ratios and higher-order terms, thus not relying on particular ordering of the outcomes (Shpitser, 2023). In order to identify the joint distribution, equality (1) shows that we only need to identify the itemwise propensity scores and the sequential odds ratio functions. These quantities are not identified without model assumptions on the missingness mechanism. The itemwise propensity score \( p(r_i|y, r_\le i = 1) \) reduces to \( p(r_i|y_i, r_\le i = 1) \) under Assumption 1. For notational convenience, we let \( \pi_i(y_i) \) denote \( p(r_i = 1|y_i, r_\le i = 1) \) and \( O_i(y_i) = \{1 - \pi_i(y_i)\}/\pi_i(y_i) \) for the itemwise odds function. We write \( \eta_i, \pi_i \) and \( O_i \) for short where it causes no confusion. Under the self-censoring model, we have the following results that are useful for identification.

**Theorem 1.** Under Assumption 1, the sequential odds ratio function \( \eta_i(\cdot) \) for \( i = 2, \ldots, p \), does not depend on \( y \) and

\[
\eta_i(r_i, r_{<i}) = \frac{p(r_{\le i}, r_{\ge i} = 1)}{E[O_i(Y_i)^{1-r_i}|r_{<i}, r_{\ge i} = 1]p(r_{<i}, r_{\ge i} = 1)}, \tag{2}
\]

Equation (2) reveals the essential role of \( O_i \) in identification of the joint distribution: once \( O_i \) is uniquely determined, \( \pi_i \) and \( \eta_i \) are both identified, which together suffice for identification of \( p(y, r) \). Besides, Equation (2) states that the sequential odds ratio \( \eta_i \) is a function of \( r_{\le i} \) but not \( y \) under the self-censoring model. Next, we connect each \( O_i \) to the observed-data distribution by

\[
E[O_i(Y_i)|y_{<i}, r = 1] = \frac{p(y_{<i}, r_i = 0, r_{\ge i} = 1)}{p(y_{<i}, r = 1)}. \tag{3}
\]

Under Assumption 1, identification of \( O_i \) is achieved as long as the solution to Equation (3) is unique, which is a Fredholm integral equation of the first kind with \( p(y_{<i}, r_i = 0, r_{\ge i} = 1), p(y_{<i}, r = 1) \) and \( p(y_i|y_{<i}, r = 1) \) obtained from the observed-data distribution and \( O_i(y_i) \) to be solved for. To identify \( O_i \), we further make the following completeness assumption.

**Assumption 2** (Completeness). \( p(y_i|y_{<i}, r = 1) \) is complete in \( y_{<i} \) for \( i = 1, \ldots, p \).

Completeness is a fundamental concept in statistics; see Lehman and Scheffe (1950). A conditional distribution \( p(u|v) \) is complete in \( v \) if \( E[g(U)|V] = 0 \) almost surely implies \( g(u) = 0 \) almost surely for any square-integrable function \( g \). Suppose \( U, V \) are discrete variables with levels \( u_k \in \{u_1, \ldots, u_K\} \) and \( v_l \in \{v_1, \ldots, v_L\} \), the completeness condition is equivalent to row full rank of matrix \( \{p(u_k|v_l)\}_{K \times L} \) with elements \( p(u_k|v_l) \). Completeness is routinely assumed in nonparametric identification problems and is viewed as a regularity condition for identification, for example, in missing data (D’Haultfœuille, 2010; Yang et al., 2019), causal inference (Miao et al., 2018; Jiang & Ding, 2021), measurement error (An & Hu, 2012), and instrumental variable identification (Newey & Powell, 2003); see Miao et al. (2022) for a review. The essential idea behind Assumption 2 is that the variability of \( Y_i \) can capture any infinitesimal variability of \( Y_i \) in the complete cases. If \( Y_i \) is fully observed, then the corresponding completeness condition on \( Y_i \) is no longer required.
We have the following identification result of full-data distribution.

**Theorem 2.** Under Assumptions 1 and 2, \( O_i(y_i) \) is identified for all \( i \) and therefore the joint distribution \( p(y, r) \) is identified.

Theorem 2 establishes identification of the full-data distribution for the self-censoring model under the completeness assumption, without imposing additional parametric model assumptions. As shown by Mohan and Pearl (2021) and Nabi et al. (2020), without restricting the observed-data distribution, identification of full-data distribution is impossible in the presence of self-censoring. However, it is possible to recover the full-data distribution by jointly harnessing additional features of the data. Assumption 2 characterizes such a feature that ensures identification under the self-censoring model. It ensures uniqueness of the solution to Equation (3), that is, identification of \( O_i \). After that, we can then identify \( \pi_i, \eta_i \) as in Equation (2) and finally identify \( p(r | y) \) as in Equation (1). Our identification strategy views \( Y_i \) as a vector of shadow variables for \( Y_i \), but in contrast to previous shadow variable approaches (D’Haultfœuille, 2010; Wang et al., 2014; Miao et al., 2023) that require a fully-observed shadow variable, here we generalize the shadow variable framework by exploiting incomplete variables \( Y_i \) as shadow variables for each missing outcome \( Y_i \) and accounting for associations between missingness of multivariate outcomes by identifying the sequential odds ratio \( \eta_i \).

### 3 ESTIMATION AND INFERENCE

#### 3.1 Parameterization

Suppose we wish to make inference about a full-data functional \( \psi \), which is defined as the unique solution of a given estimating equation \( E(m(X, Y; \psi)) = 0 \). For instance, the outcomes mean \( \psi = E(Y) \) corresponds to \( m(x, y; \psi) = y - \psi \). Having established identification of the full-data distribution, we can in principle first estimate \( p(x, y, r = 1), p(x, y, r_i = 0, r_{-i} = 1) \) and then plug them into Equation (3) to solve for the itemwise odds function \( O_i \), and finally obtain the joint distribution \( p(x, y, r) \) and any functional of interest. Although the first step can be achieved by standard nonparametric estimation, the convergence rate is slow particularly when multivariate outcomes are concerned, and the situation becomes worse if additional fully-observed covariates \( X \) must be taken into account. Besides, it is challenging to solve Equation (3) due to the ill-posedness of the Fredholm equations of the first kind, which further complicates statistical inference.

Therefore, we consider a parameterization of the joint distribution and develop semiparametric estimators that only require the working models to be partially correct. The full-data distribution can be expressed by the odds ratio factorization,

\[
p(y, r | x) = \frac{OR(y, r | x)p(r | x = 1, x)p(r | x, y)}{\sum_{r' \in R} E[OR(Y, r' | x)x = 1, x]p(r' | x, y)},
\]

which is widely used in missing data problems, see Chen (2007), Kim and Yu (2011), Tchetgen Tchetgen et al. (2018), and Malinsky et al. (2022), for examples. Here, \( y = y_0 \) is a reference value in the support of \( y \) chosen by the analyst and

\[
OR(y, r | x) = \frac{p(r | x, y)p(r = 1 | x, y_0)}{p(r | x, y_0)p(r = 1 | x, y)},
\]

which is the odds ratio function between the vectors \( Y \) and \( R \) conditional on \( X \), capturing the dependence of the missingness indicators on the outcomes owing to a shift of \( Y \) from the reference value \( y_0 \). Equation (4) reveals a congenial specification of the joint distribution with three variationally independent components: the baseline outcome distribution \( p(y | x, r = 1) \); the odds ratio function \( OR(y, r | x) \), which can be factorized by a product of itemwise odds ratio \( \Gamma_i(x, y_i) = O_i(x, y_i)/O_i(x, y_{0i}) \),

\[
OR(y, r | x) = \prod_{i=1}^p \Gamma_i(x, y_i)^{1-r_i},
\]

where \( y_{0i} \) denotes the \( i \)th entry of \( y_0 \); the baseline propensity score \( p(r | x, y_0) \) for each missingness pattern \( r \). We consider statistical inference of \( \psi \) under partially correct specification of these three models. In the following, we let \( \Pi_r(x, y) = p(r | x, y) \) denote the propensity score for missingness pattern \( r \). We use \( \hat{E} \) for the empirical mean.

#### 3.2 Inverse probability weighted estimator

We first consider an IPW approach that entails a parametric working model for the propensity score \( \Pi_r(x, y) \) for each missingness pattern \( r \). We model \( \Pi_r(x, y) \) by specifying parametric working models for the odds ratio function \( OR(y, r | x; \gamma) \) and the baseline propensity score \( \Pi_r(x, y_0; \alpha) \), which is equivalent to modeling each \( \Gamma_i(x, y_i; \gamma_i), \pi_i(x, y_0; \alpha_i1) \) (or \( O_i(x, y_0; \alpha_i1) \) equivalently), \( \eta_i(r_i; r_{-i}, x; \alpha_2) \) and evaluating \( \Pi_r(x, y) \) according to Equation (1). Let \( y = (y_1, ..., y_p) \), \( \alpha_1 = (\alpha_{11}, ..., \alpha_{1p}) \), \( \alpha_2 = (\alpha_{22}, ..., \alpha_{2p}) \), and \( \alpha = (\alpha_1, \alpha_2) \) denote the parameters for the working models. The itemwise propensity score is determined by the itemwise odds ratio function \( \Gamma_i(x, y_i; \gamma_i) \).
and itemwise baseline odds function \( O_i(x, y_{ij}; \alpha_{i1}) \) by \( \pi_i(x, y; \alpha_{i1}, \gamma_{ij}) = \{1 + O_i(x, y_{ij}; \alpha_{i1})G_i(x, y_{ij}; \gamma_{ij})\}^{-1} \).

The estimation of \((\alpha, \gamma)\) is motivated by the fact that \( E\{1(R = 1)\Pi_i / \Pi_i - 1(R = r)|X\} = 0\), where \( 1(\cdot) \) denotes the indicator function. The first function is a consequence of self-censoring and the second equation echoes the definition of \( \Pi_i \) by equality (1). We estimate \((\alpha, \gamma)\) by solving the observed data estimating equations

\[
0 = \hat{E}\left\{ \frac{1(R = 1)}{\pi_i(\hat{\alpha}_{1i}, \hat{\gamma}_i)} - 1(R_{-i} = 1) \right\} \cdot g_i,
\]

\[
0 = \hat{E}\left\{ \left( 1(R = 1) \prod_{j: r_j = 0} O_j(\hat{\alpha}_{1j}, \hat{\gamma}_j) \prod_{i=2}^p \eta_i(r_i, r_{-i}, X; \hat{\alpha}_{2i}) \right) - 1(R = r) \right\} \cdot h_i,
\]

for all \( i \), where \( g_i(x, Y_{-i}) \) and \( h_i(x) \) are user-specified vector functions of the same dimension as \((\alpha_{1i}, \gamma_{ij})^T \) and \( \alpha_{2i} \), respectively. Specifically, if one uses a logistic model for itemwise odds \( O_i(x, y_{ij}; \alpha_{i1}, \gamma_{ij}) = \exp(\alpha_{i1} x + \gamma_{ij} y) \), and a bilinear model (Chen, 2004) for the sequential odds ratio function \( \eta_i(r_i, r_{-i}, x; \alpha_{2i}) = \exp(\alpha_{2i} x) \) for \( r_i = 0 \) and \( r_{-i} \neq 1 \), then one natural choice for \( h_i \) is \( \delta \log(\eta_i(r_i, r_{-i}, x; \alpha_{2i})) / \delta \alpha_{2i} = x \). In Equation (6), \( Y_{-i} \) is observed for \( R_{-i} = 1 \) and thus used as a proxy for \( Y_{-i} \) to capture the variation of \( Y_i \) that is not available from the pattern \( R_{-i} = 1 \). One may choose \( g_i(x, Y_{-i}) = \delta \log O_i(x, y_{-i}, \alpha_{i1}, y_{i1}) / \delta \alpha_{i1} = (x^T, y_{-i}) \) where \( y_{-i} \) is the average of outcomes in \( Y_{-i} \).

Let \( (\hat{\alpha}_{pw}, \hat{\gamma}_{pw}) \) denote the nuisance estimators obtained from Equations (6) and (7). To estimate \( \psi \), we solve the following estimating equation,

\[
\hat{E}\left\{ \frac{1(R = 1)}{\Pi_1(\hat{\alpha}_{pw}, \hat{\gamma}_{pw})} m(X, Y; \hat{\psi}_{pw}) \right\} = 0,
\]

where the full-data estimating equation is evaluated in the complete cases and the inverse propensity score weighting removes the selection bias due to missing data.

### 3.3 Outcome regression/imputation-based estimator

Alternatively, we can estimate \( \psi \) by imputing the missing values. A commonly-used approach for imputation is the exponential tilting or Tukey’s representation (Kim & Yu, 2011; Franks et al., 2020): for each missingness pattern \( R = r' \),

\[
p(y|x, r') = \frac{OR(y, r'|x)p(y|x, r = 1)}{E[OR(Y, r'|x)|x, r = 1]}.
\]

The imputation entails a baseline outcome model \( p(y|x, r = 1; \beta) \) for the complete cases and an odds ratio model \( OR(Y, r|x; \gamma) \). We solve the following equations to obtain the nuisance estimators \((\beta, \gamma)\).

\[
0 = \hat{E}\left\{ \frac{1(R = 1)}{\Pi_i(\hat{\alpha}_{1i}, \hat{\gamma}_i)} - 1(R_{-i} = 1) \right\} \cdot \{ g_i - \hat{E}(g_i|X, R_i = 0, R_{-i} = 1; \hat{\beta}, \hat{\gamma}_i) \},
\]

for \( i = 1, \ldots, p \). The conditional expectation \( \hat{E}(\cdot|R_i = 0, R_{-i} = 1) \) is evaluated under Equation (9) and \( g_i(x, y_{-i}) \) is a user-specified function of the same dimension as \( \gamma_{ij} \). Estimation of \( \beta \) only involves complete cases. Estimation of the odds ratio parameter \( \gamma \) in Equation (10) is motivated by the fact that \( E(1(R_i = 0, R_{-i} = 1) \cdot \log(OR(Y, r_i|x; \gamma) | R_i = 0, R_{-i} = 1, X)) = 0 \) for any function \( l(x, y) \). Because \( Y_i \) is missing for \( R_i = 0 \), we replace \( l(x, y) \) with \( g(x, y_{-i}) \) where \( Y_{-i} \) is observed for \( R_{-i} = 1 \) and used as shadow variables for \( Y_i \).

Let \( (\hat{\beta}_{reg}, \hat{\gamma}_{reg}) \) denote the nuisance estimators obtained from Equations (10) and (11). An outcome regression/imputation-based estimator for \( \psi \) is given by solving

\[
\hat{E}\left\{ 1(R = 1) \cdot m(\hat{\psi}_{reg}) + \sum_{r=1}^R 1(R=r) \cdot \hat{E}\left\{ m(\hat{\psi}_{reg})|X, Y_r \right\}, \quad R=r; \hat{\beta}_{reg}, \hat{\gamma}_{reg} \right\} = 0,
\]

where \( \hat{E}(\cdot|X, Y_r, R = r; \hat{\beta}, \hat{\gamma}) \) is evaluated according to Equation (9) to impute the missing values for pattern \( R = r \).

### 3.4 Doubly robust estimator

However, consistency of the IPW and regression/imputation-based estimators is no longer guaranteed if any of the required working models is incorrect. Therefore, we construct an estimator that combines both approaches and achieves double robustness against misspecification of the working models. We estimate the nuisance parameters by solving Equations (7) and (10), and

\[
\hat{E}\left\{ \left( \frac{1(R = 1)}{\Pi_i(\hat{\alpha}_{1i}, \hat{\gamma}_i)} - 1(R_{-i} = 1) \right) \cdot \{ g_i - \hat{E}(g_i|X, R_i = 0, R_{-i} = 1; \hat{\beta}, \hat{\gamma}_i) \} \right\} = 0.
\]
for all $i$ together and let $(\hat{\alpha}_{dr}, \hat{\beta}_{dr}, \hat{\gamma}_{dr})$ denote the nuisance estimators. The doubly robust estimator of $\psi$ is given by the solution to Equation (14).

\[
E \left[ \frac{1(R = 1)}{\Pi_1(\hat{\alpha}_{dr}, \hat{\gamma}_{dr})} m(\hat{\psi}_{dr}) + \sum_{r \neq 1} \frac{1(R = r)}{\Pi_1(\hat{\alpha}_{dr}, \hat{\gamma}_{dr})} \right] = 0.
\]

Equations (7) and (10) have been used in the IPW and outcome regression/imputation-based approaches for estimation of $\alpha$ and $\beta$, respectively, while in contrast we use a different equation (13) for estimation of $\gamma$ and Equation (14) for estimation of $\psi$. Equations (13) and (14) have an augmented IPW form, where an augmentation term involving the outcome regression is included to correct the bias of the IPW estimation when the baseline propensity score model is incorrect. The resultant estimators $\hat{\gamma}_{dr}$ and $\hat{\psi}_{dr}$ are in fact doubly robust.

3.5 | Theoretical results

Suppose $\mathcal{M}_{r,j}$ represents the model where $\pi_i(x, y_0; \alpha_{1j})$ and $\Gamma_i(x, y; \gamma_j)$ are correctly specified; and $\mathcal{M}_{y,j}$ represents $p(y|x, r = 1; \beta)$ and $\Gamma_i(x, y; \gamma_j)$ are correctly specified. The following theorem summarizes properties of the odds ratio estimators.

**Theorem 3.** Under Assumptions 1 and 2, and the regularity conditions in Theorems 2.6 and 3.4 of Newey and McFadden (1994):

(i) $\hat{\gamma}_{ipw}$ is consistent and asymptotically normal in model $\mathcal{M}_{r,j}$;
(ii) $\hat{\gamma}_{reg}$ is consistent and asymptotically normal in model $\mathcal{M}_{y,j}$;
(iii) $\hat{\psi}_{dr}$ is consistent and asymptotically normal in the union model $\mathcal{M}_{r,i} \cup \mathcal{M}_{y,j}$.

Let $\mathcal{M}_r = \cap_{l=1}^p \mathcal{M}_{r,l} \cap (\cap_{l=2}^p \eta_l(r, r_2, x; \alpha_2))$ is correct), that is, $\Pi_l(x, y_0; \alpha)$ and OR($y, r|x; \gamma$) are correctly specified; and $\mathcal{M}_y = \cap_{l=1}^p \mathcal{M}_{y,l}$, that is, $p(y|x, r = 1; \beta)$ and OR($y, r|x; \gamma$) are correctly specified. The following theorem summarizes properties of the estimators of $\psi$.

**Theorem 4.** Under Assumptions 1 and 2, and the regularity conditions in Theorems 2.6 and 3.4 of Newey and McFadden (1994):

(i) $\hat{\gamma}_{ipw}$ is consistent and asymptotically normal in model $\mathcal{M}_{r,j}$;
(ii) $\hat{\gamma}_{reg}$ is consistent and asymptotically normal in model $\mathcal{M}_{y,j}$;
(iii) $\hat{\psi}_{dr}$ is consistent and asymptotically normal in the union model $\mathcal{M}_{r,i} \cup \mathcal{M}_{y,j}$.

Doubly robust methods have been promoted for missing data analysis, causal inference and many other coarsened data problems, see Seaman and Vansteelandt (2018) for a review. Theorems 3 and 4 exhibit the double robustness of $\hat{\gamma}_{dr}$ and $\hat{\psi}_{dr}$. The consistency of $\hat{\gamma}_{dr}$ requires correct specification of the itemwise odds ratio $\Gamma_i(x, y; \gamma_j)$ and of at least one of the baseline models $\{\pi_i(x, y_0; \alpha_{1j}), p(y|x, r = 1; \beta)\}$. The consistency of $\hat{\psi}_{dr}$ requires stronger conditions—correct specification of OR($y, r|x; \gamma$) that is, all the itemwise odds ratio models and correct specification of at least one of $\{\Pi_l(x, y_0, \alpha), p(y|x, r = 1; \beta)\}$. The doubly robust estimators offer one more chance to correct bias due to partial model misspecification; although, they will generally also be biased if both baseline models are incorrect (Kang & Schafer, 2007) or the odds ratio function is incorrect. Following the general theory for estimation equations (Newey & McFadden, 1994), variance estimation of the estimators can be obtained and confidence intervals can be constructed based on the normal approximation. One can also implement the bootstrap method.

4 | SENSITIVITY ANALYSIS

The self-censoring assumption is stringent in some cases and part of them are not empirically testable. Therefore, it is crucial to assess the deviation from the inferences in a given analysis, when the self-censoring assumption is violated. We introduce a global sensitivity analysis approach under a class of models anchored around the self-censoring assumption.

**Assumption 3** (Mixed-censoring). $R_i \perp Y_{s(i)} | (X, Y_i, R_{s(i)})$ and $R_i \perp Y_{s(i)} | (X, Y_i, R_{s(i)})$, for $i = 1, \ldots, p$, where $s(i) \subset \{1, \ldots, p\} \setminus \{i\}$ and $s(i) = \{1, \ldots, p\} \setminus \{s(i) \cup \{i\} \}$. Assumption 3 allows for a relaxation of the self-censoring that a missingness indicator $R_i$ may not be conditional independent of a subset of outcomes $Y_{s(i)}$. Such conditional independence can be characterized by
the existence of directed edges from \( Y_0(i) \) to \( R_i \) by the local Markov property of chain graph, see Figure 2 for example.

Recall that the odds ratio factorization (1) is valid under arbitrary missingness mechanism. Under Assumption 3, \( p(r_i|x, y, r_{-i}) = 1 \) reduces to \( p(r_i|x, y, y_0(i), r_{-i}) = 1 \). With a little abuse of notations, we let \( O_i(x, y_i, y_0(i)) = \{1 - \pi_i(x, y_i, y_0(i))\}/\pi_i(x, y_i, y_0(i)) \) denote the itemwise odds function, where \( \pi_i(x, y_i, y_0(i)) = p(r_i = 1|x, y_i, y_0(i), r_{-i} = 1) \). We specify each itemwise odds \( O_i \) with an additive structure \( O_i(x, y_i, y_0(i)) = \exp(\gamma_i(x, y_i) + \theta_i(x, y_0(i)), \) where \( \gamma_i(x, y_i) \) is an arbitrary unknown function, and \( \theta_i(x, y_0(i)) \) is a specified functional form in a sensitivity analysis (Robins et al., 2000) measuring the dependency of \( R_i \) on \( Y_0(i) \). For specified \( \theta_i \), the unknown function \( \gamma_i(x, y_i) \) is identified under Assumptions 2 and 3.

To recover the missingness mechanism, we further need the sequential odds ratio \( \eta_i \) which characterizes the dependence among \( R_i \). Under Assumption 3, \( \eta_i(\cdot) \) is a function of \( (r_{i\neq i}, x, y_0(i)) \) where \( e(i) = \{(i) \cup s(i)\} \cap \{U<j \leq i \cup s(j)\} \). For example, in Figure 2, \( e(2) = \{1\} \) and \( e(3) = \{2\} \). Further under Assumption 2, if \( e(i) = \emptyset \), then \( \eta_i \) can be identified by \( O_i \) and observed-data distribution. Therefore, one should specify the known functional form of the non-negative function \( \eta_i(r_{i\neq i}, x, y_0(i)) \) if \( e(i) \neq \emptyset \), while model \( \eta_i(r_{i\neq i}, x) \) with unknown parameters if \( e(i) = \emptyset \). From all above, the joint distribution is identified when the sensitivity parameters \( \{(s(i), \theta_i, \eta_i)\} \) capturing the deviation away from self-censoring are specified.

**Theorem 5.** Under Assumptions 2 and 3, \( p(x, y, r) \) is identified for specified \( \{(s(i), \theta_i, \eta_i)\} \).

We describe three estimating approaches of \( \hat{\psi} \) incorporating a non-null \( \{(s(i), \theta_i, \eta_i)\} \) in Web Appendix B. The sensitivity analysis reports on the range of \( \{(\hat{\psi}_{pw}, \hat{\psi}_{reg}, \hat{\psi}_{dr})\} \) within the range of \( \{(s(i), \theta_i, \eta_i)\} \). Our approach is interpretable and flexible, but the model specification is complicated when the cardinality of \( s(i) \) is large. However, the aim of sensitivity analysis is to check certain degree of violation of self-censoring, so the index set \( s(i) \) should be carefully chosen with specialized knowledge or auxiliary information. For multivariate nonignorable nonmonotone missing data, Tompsett et al. (2018) proposed a sensitivity analysis procedure that imputes each outcome with a model adjusting for the other outcomes and missingness indicators. Scharfstein et al. (2022) developed a sensitivity analysis method in longitudinal studies for binary outcomes anchored at the assumption introduced by Robins (1997). Graphically, their approaches treat the effects of self-censoring edges as sensitivity parameters, while ours treat the effects of the other edges as sensitivity parameters.

## 5 SIMULATION STUDY

We evaluate the performance of the proposed estimation methods via simulations. We generate a covariate \( X \sim \text{Uniform}(-1, 1) \), three outcomes \( (Y_1, Y_2, Y_3) \) and corresponding \( (R_1, R_2, R_3) \) from the following model:

\[
R_i|X, Y_i, R_{-i} = 1 \sim \text{Bernoulli}(a_i^T u(X) + c_i Y_i), i = 1, 2, 3; \quad Y_i|X, R_i = 1 \sim N(b_i^T v(X), \Sigma) \; \quad \eta_1(r_2, r_1, x) = \exp\{(1 - r_1)(1 - r_2)x/10\}; \quad \eta_2(r_3, r_12, x) = \exp\{(1 - r_3)(1 - r_12)x/10\};
\]

where \( \Sigma \) is a matrix with 1 on the diagonal and 0.5 elsewhere. Four scenarios with different choices of \( u(x), v(x) \) and an additional missing at random (MAR) scenario with \( c_1 = c_2 = c_3 = 0 \) are considered. The detailed parameter settings are described in Web Table F2.

We are interested in estimating \( \psi = E(Y_1) \). For estimation, we specify a multivariate normal distribution \( N((1, x)\beta, \Sigma) \) for the baseline outcome distribution \( p(y|x, r = 1) \) and a bilinear model for the odds ratio function \( \exp\{\sum_{i=1}^3 (1 - r_i)\gamma_i y_i\} \) for each baseline itemwise odds \( O_i(x, y_0(i)) \) and a bilinear model for the sequential odds ratios \( \eta_2 = \exp(\alpha_{21}(1 - r_1)(1 - r_2)x), \eta_3 = \exp(\alpha_{22}(1 - r_3)(1 - r_12)x) \). We compare the proposed estimators to a benchmark doubly robust estimator for missing at random (MARDR). We also compute the sandwich variance estimator and the Wald-type confidence interval.

We replicate 10,000 simulations for each scenario with sample size 3,000 and summarize the results for estimation bias of the outcome mean \( \psi \) and odds ratio parameter \( \gamma_1 \) in Figure 3, Table 1 shows mean of the bias of the point estimators, median of the standard error estimates, and coverage rate of the 95% confidence interval. In scenario TT, all working models are correct, then the three proposed estimators show little bias and the coverage rate of the 95% confidence interval is close to 0.95. In

![Figure 2](image-url)
TABLE 1 Results from simulation study described in Section 5: mean of the bias of the point estimates (×100), median of the sandwich variance-based standard error estimates, and coverage rate (×100) of the 95% Wald confidence intervals.

| Outcome mean (ψ) | DR | IPW | REG | MARDR | Odds ratio parameter (γ_1) | DR | IPW | REG |
|------------------|----|-----|-----|-------|---------------------------|----|-----|-----|
| TT | -0.089 | 0.092 | 0.011 | -7.137 | -0.178 | -0.114 | 0.125 |
| TT Mean | -0.120 | 0.074 | 4.608 | -34.37 | 0.636 | 0.312 | -12.10 |
| FF of bias | -0.410 | 24.40 | -0.061 | -6.631 | -0.546 | 13.57 | 0.064 |
| MAR | -0.004 | 0.033 | 0.000 | -0.003 | 0.022 | -0.059 | 0.041 |
| TT Median | 0.036 | 0.052 | 0.035 | 0.022 | 0.111 | 0.112 | 0.112 |
| TT SE | 0.062 | 0.081 | 0.065 | 0.074 | 0.085 | 0.086 | 0.045 |
| MAR of SE | 0.096 | 0.224 | 0.095 | 0.042 | 0.175 | 0.215 | 0.172 |
| TT Coverage | 96.25 | 96.72 | 95.24 | 13.13 | 95.39 | 95.16 | 95.21 |
| TT rate | 96.32 | 96.87 | 88.17 | 16.03 | 94.64 | 95.11 | 24.11 |
| FF | 85.55 | 89.48 | 79.37 | 19.29 | 87.34 | 89.95 | 86.86 |
| MAR | 96.21 | 97.16 | 95.34 | 95.58 | 95.55 | 95.39 | 95.14 |

Abbreviations: DR, doubly robust estimator; IPW, inverse probability weight; MARDR, doubly robust estimator for missing at random; REG, regression; SE, standard error.

scenarios TF, Π_r(x, y_0; α) and OR(y, r|x; γ) are correct but p(y|x, r = 1; β) is incorrect, then the IPW and DR estimators show little bias with fine coverage rates, while the REG estimator has non-negligible bias with an undercoverage confidence interval. In scenario FT, Π_r(x, y_0; α) is incorrect but p(y|x, r = 1; β) and OR(y, r|x; γ) are correct, then the IPW estimator has non-negligible bias, large variance and a coverage rate well below 0.95. While the REG and DR estimators have little bias with coverage rates close to 0.95. However, when both baseline working models are misspecified (FF), all three proposed estimators are biased. As expected, the MARDR has large bias in all the above four scenarios with a low coverage rate. Under the MAR scenario, all estimators are unbiased, while the proposed estimators for ψ show larger variance than MARDR because they do not make use of the priori information that the missingness is at random. In summary, we recommend the DR estimator for multivariate missing data subject to self-censoring. In Web Appendix B, we also assess the robustness of the estimators when the self-censoring assumption is violated. The results show that the bias is relatively small as the sensitivity parameters vary within a moderate range, but become non-negligible when the model departs far from self-censoring.

6 RE-ANALYSIS OF HIV DATA

We analyze a dataset extracted from an observational study of HIV-positive mothers in Botswana. It is of interest to evaluate the association between maternal HAART in pregnancy and preterm delivery adjusted for CD4 lymphocyte cell count. Details of the study are described in Chen et al. (2012). We focus on n = 2341 HIV-positive women with maternal hypertension. Our analysis includes three binary outcomes, namely an indicator of preterm delivery Y_1 (15.5% missing), an indicator of whether CD4 cell count is less than 200 Y_2 (41.0% missing), and the missingness is nonmonotone. Our primary interest is the risk difference of continuing HAART on preterm delivery after adjustment for two CD4 cell count levels. Let RD_1, RD_2 denote the risk differences in the low and normal CD4^+ cell count groups, respectively.

Chen et al. (2012) mentioned that “Several variables in the obstetric records were self-reported, and some information was missing.” We suspect the data suffer from nonignorable missing due to social stigma. Therefore, we apply the self-censoring model and implement the proposed methods for estimation. We use multinomial distributions for the baseline propensity score Π_r(y_0) and the baseline outcome models p(y|r = 1) for binary outcomes, and model the odds ratio with OR(y, r) = exp{∑_{i=1}^3(1 – r_i)γ_iy_i}, where each γ_i characterizes the extent of nonignorable missingness of each outcome. We compare our methods to the augmented IPW estimator under NSC model from Malinsky et al. (2022) and multivariate imputation by chained equations (MICE) from Van Buuren and Groothuis-Oudshoorn (2011), where each incomplete
TABLE 2  Point estimate and 95% confidence interval of odds ratio parameters.

| Method | \( \gamma_1 \)         | \( \gamma_2 \)         | \( \gamma_3 \)          |
|--------|------------------------|------------------------|-------------------------|
| IPW    | -0.10                  | -3.32                  | -1.03                   |
| REG    | 0.01                   | -5.32                  | -1.19                   |
| DR     | 0.06                   | -5.10                  | -1.06                   |

Abbreviations: DR, doubly robust estimator; IPW, inverse probability weight; REG, regression.

TABLE 3  Point estimate and 95% confidence interval of risk differences.

| Method | \( RD_1 \)         | \( RD_2 \)         |
|--------|-------------------|-------------------|
| IPW    | 0.292 (-0.042, 0.649) | 0.157 (-0.013, 0.305) |
| REG    | 0.266 (-0.040, 0.586) | 0.153 (-0.009, 0.301) |
| DR     | 0.267 (-0.042, 0.597) | 0.153 (-0.011, 0.302) |
| NSC    | 0.289 (-0.046, 0.699) | 0.184 (0.046, 0.315) |
| MICE   | 0.097 (-0.028, 0.427) | 0.060 (-0.148, 0.174) |

Abbreviations: DR, doubly robust estimator; IPW, inverse probability weight; MICE, Multivariate imputation by chained equation; NSC, no self-censoring; REG, regression.

FIGURE 3  Bias of estimators. The first figure represents bias of the outcome mean estimators and the second figure shows bias of the odds ratio estimators. TT stands for correct working model for baseline propensity score but incorrect working model for baseline outcome model, and TT, FT, FF are similarly defined, respectively.

outcome is imputed with a univariate logistic regression model assuming MAR and 10 imputed datasets are pooled to obtain the estimates. The 95% confidence intervals are obtained with bootstrap.

Table 2 shows the estimates of the odds ratio parameters \( \gamma \). The point estimation of \( \gamma_1 \) is close to zero, suggesting that the missingness of preterm delivery is prone to be independent of its value. However, the estimates of \( \gamma_2 \) and \( \gamma_3 \) are significantly negative. This is evidence for nonignorable missingness and suggests that mothers with low CD4+ count or receiving HAART are more likely to respond. One possible explanation is these mothers pay more attention to their own health status and more willing to cooperate with data collection. Table 3 reports the estimates of the risk differences \( RD_1 \) and \( RD_2 \). The three estimators based on the self-censoring model exhibit that continuing HAART contributes to this risk of preterm delivery for mothers with HIV-positive and maternal hypertension. We also compare our estimates with the estimator based on NSC model, and the conclusion is consistent in the way that both models detect significant risk differences for the two groups. In contrast, the multiple imputation method substantially underestimates the two risk differences, because it does not account for nonignorable missingness. To check the robustness of the above results, we conduct a sensitivity analysis with \( OR(y, r) = \exp\{(1 - r_1)\delta y_2 + \sum_{i=1}^{3} (1 - r_i)\gamma_i y_i\} \), where \( \delta \), ranging from -2 to 2, is the sensitivity parameter encoding the dependence of preterm delivery nonresponse on the CD4+ level. The curves in Figure 4 display the sensitivity of the estimation as one deviates away from the benchmark assumption \( \delta = 0 \). The estimated risk differences are always positive, which is evidence supporting the results from benchmark analysis. If the missingness of preterm delivery is more likely to depend on low CD4+ cell count, then Figure 4a shows that the risk difference \( RD_1 \) might be even larger, and Figure 4b depicts that \( RD_2 \) retains (IPW and REG estimators) or slightly decreases (DR estimator).

7 | DISCUSSION

The self-censoring model imposes restrictions on the observed-data distribution, thus not nonparametric saturated. As suggested by D’Hautefoeuille (2010), part of condition (ii) in Assumption 1 is empirically falsifiable.
using observed data by testing the existence of a solution to certain equations, see Web Appendix A. Assumption 2 implicitly requires the number of levels of $Y_i$ is no smaller than that of $Y_1$. The completeness can be assessed in specific models, that is, nonsingularity of the covariance matrix in the joint normal model (Web Appendix D). However, it is untestable without restrictions on the distribution (Canay et al., 2013).

Our estimation methods require sufficient number of complete-case observations, however, with multiple missing outcomes, the number of complete cases may be small. In the meanwhile, the joint distribution is more likely to be incorrectly specified. One can incorporate certain Markov dependence structure to simplify the model specification and estimation. It is of interest to develop the semiparametric efficiency bound and the efficient estimator. Yet, due to the complexity of the multivariate self-censoring model, the closed forms for the efficient influence function and efficient estimator are currently not available. Moreover, the efficient estimator involves complex features of the observed-data distribution which are difficult to model correctly, and thus the potential prize of implementing the efficient estimator may not always be worth to chase (Stephens et al., 2014).

In addition to self-censoring, the missingness of an outcome may depend on others in practice. In Web Appendix C, we extend the self-censoring model to a blockwise self-censoring model and a composite model that admits mixed-censoring mechanisms. We establish identifying conditions for such models and plan to study their statistical inference in the future.

**ACKNOWLEDGMENTS**

We are grateful for valuable comments from the editor, the associate editor and one reviewer. This research is supported by the National Key R&D Program (2022YFA1008100), National Natural Science Foundation of China (12071015), Beijing Natural Science Foundation (Z1900001), ONR grant (N00014-21-1-2820), NSF grant (2040804, CAREER 1942239), and NIH grant (ROI AI1127271-01A1).

**DATA AVAILABILITY STATEMENT**

The data that support the findings in this paper are available in the Supporting information.

**ORCID**

Yilin Li https://orcid.org/0000-0001-7124-9186

**REFERENCES**

An, Y. & Hu, Y. (2012) Well-posedness of measurement error models for self-reported data. *Journal of Econometrics*, 168, 259–269.

Brown, C.H. (1990) Protecting against nonrandomly missing data in longitudinal studies. *Biometrics*, 46, 143–155.

Canay, I.A., Santos, A. & Shaikh, A.M. (2013) On the testability of identification in some nonparametric models with endogeneity. *Econometrica*, 81, 2535–2559.

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

Chen, H.Y. (2007) A semiparametric odds ratio model for measuring association. *Biometrics*, 63, 413–421.

Chen, H.Y. (2010) Compatibility of conditionally specified models. *Statistics & probability letters*, 80, 670–677.

Chen, J.Y., Ribaudo, H.J., Souda, S., Parekh, N., Ogwu, A., Lockman, S. et al. (2012) Highly active antiretroviral therapy and adverse birth outcomes among hiv-infected women in Botswana. *The Journal of Infectious Diseases*, 206, 1695–1705.

D'Haultfoeuille, X. (2010) A new instrumental method for dealing with endogenous selection. *Journal of Econometrics*, 154, 1–15.

Du, H., Enders, C., Keller, B.T., Bradbury, T.N. & Karney, B.R. (2022) A bayesian latent variable selection model for nonignorable missingness. *Multivariate Behavioral Research*, 57, 478–512.

Fay, R.E. (1986) Causal models for patterns of nonresponse. *Journal of the American Statistical Association*, 81, 354–365.

Franks, A.M., D’Amour, A. & Feller, A. (2020) Flexible sensitivity analysis for observational studies without observable implications. *Journal of the American Statistical Association*, 115, 1730–1746.

Jiang, Z. & Ding, P. (2021) Identification of causal effects within principal strata using auxiliary variables. *Statistical Science*, 36, 493–508.
Sadinle, M. & Reiter, J.P. (2017) Itemwise conditionally independent nonresponse modelling for incomplete multivariate data. *Biometrika*, 104, 207–220.

Scharfein, D.O., Steingrimsson, J., McDermott, A., Wang, C., Ray, S., Campbell, A., Nunes, E. & Matthews, A. (2022) Global sensitivity analysis of randomized trials with nonmonotone missing binary outcomes: application to studies of substance use disorders. *Biometrika*, 78, 649–659.

Seaman, S.R. & Vansteelandt, S. (2018) Introduction to double robust methods for incomplete data. *Statistical Science*, 33, 184.

Shpitser, I. (2016) Consistent estimation of functions of data missing non-monotonically and not at random. *Advances in Neural Information Processing Systems*, 29, 3144–3152.

Shpitser, I. (2023) The Lauritzen–Chen likelihood for graphical models. *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, vol. 206. PMLR. pp. 4181–4195.

Stephens, A., Tchetgen Tchetgen, E. & De Gruttola, V. (2014) Locally efficient estimation of marginal treatment effects when outcomes are correlated: is the prize worth the chase? *The International Journal of Biostatistics*, 10, 59–75.

Sun, B., Liu, L., Miao, W., Wirth, K., Robins, J. & Tchetgen Tchetgen, E. (2018) Semiparametric estimation with data missing not at random using an instrumental variable. *Statistica Sinica*, 28, 1965–1983.

Tchetgen Tchetgen, E., Wang, L. & Sun, B. (2018) Discrete choice models for nonmonotone nonignorable missing data: identification and inference. *Statistica Sinica*, 28, 2069–2088.

Tompsett, D.M., Leacy, F., Moreno-Betancur, M., Heron, J. & White, I.R. (2018) On the use of the not-at-random fully conditional specification (narfc) procedure in practice. *Statistics in Medicine*, 37, 2338–2353.

Van Buuren, S. & Groothuis-Oudshoorn, K. (2011) mice: multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45, 1–67.

Wang, S., Shao, J. & Kim, J.K. (2014) An instrumental variable approach for identification and estimation with nonignorable nonresponse. *Statistica Sinica*, 24, 1097–1116.

Yang, S., Wang, L. & Ding, P. (2019) Causal inference with confounders missing not at random. *Biometrika*, 106, 875–888.

Zhou, Y., Little, R.J. & Kalbfleisch, J.D. (2010) Block-conditional missing at random models for missing data. *Statistical Science*, 25, 517–532.

**SUPPORTING INFORMATION**

Web Appendices referenced in Sections 2–7 are available with this paper at Biometrics website on Wiley Online Library. R codes to replicate the simulations and application results are also provided.

**How to cite this article:** Li, Y., Miao, W., Shpitser, I. & Tchetgen Tchetgen, E.J. (2023) A self-censoring model for multivariate nonignorable nonmonotone missing data. *Biometrika*, 79, 3203–3214. https://doi.org/10.1111/biom.13916