Received: 27 August 2018 | Accepted: 25 July 2019

DOI: 10.1111/biom.13131

BIOMETRIC METHODOLOGY

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

Improving estimation efficiency for regression with MNAR covariates

Menglu Che[1] | **Peisong Han**[2] | **Jerald F. Lawless**[1]

[1]Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada

[2]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan

Correspondence
Peisong Han, Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan 48103.
Email: peisong@umich.edu

Abstract

For regression with covariates missing not at random where the missingness depends on the missing covariate values, complete-case (CC) analysis leads to consistent estimation when the missingness is independent of the response given all covariates, but it may not have the desired level of efficiency. We propose a general empirical likelihood framework to improve estimation efficiency over the CC analysis. We expand on methods in Bartlett *et al.* (2014, *Biostatistics* **15**, 719–730) and Xie and Zhang (2017, *Int J Biostat* **13**, 1–20) that improve efficiency by modeling the missingness probability conditional on the response and fully observed covariates by allowing the possibility of modeling other data distribution-related quantities. We also give guidelines on what quantities to model and demonstrate that our proposal has the potential to yield smaller biases than existing methods when the missingness probability model is incorrect. Simulation studies are presented, as well as an application to data collected from the US National Health and Nutrition Examination Survey.

KEYWORDS

complete-case analysis, empirical likelihood, estimating equations, missing covariates, missing not at random

1 | INTRODUCTION

Regression analysis is often complicated by the presence of missing data. Handling missing data inappropriately can lead to biased estimation and/or loss of efficiency. The most commonly used assumption about the missingness mechanism is missing-at-random (MAR), where the missingness depends on the observed data but not on the missing data. There is a rich collection of effective methods dealing with MAR data, including multiple imputation (Rubin, 1987), inverse probability weighting (Horvitz and Thompson, 1952), augmented inverse probability weighting (Robins *et al.*, 1994), and other likelihood-based methods (Little and Rubin, 2002). However, in many settings, the assumption of MAR is too strong, and the missingness does depend on the missing data even conditional on the observed data. Developing general methods dealing with such missing-not-at-random (MNAR) data is very challenging due to model identifiability issues. See, for example, Rotnitzky and Robins (1997), Ibrahim *et al.* (1999), Wang *et al.* (2014), Miao and Tchetgen Tchetgen (2016), and Han (2018), for some relevant discussions.

In this article, we consider regression analysis with MNAR covariates where the missingness is assumed to be independent of the response given by all covariates of interest. This is a practically important setting, especially when the covariates are measured at the beginning of the study but the response is measured at a later time point. In this case, it is natural and logical to assume that the missingness of covariates does not depend on the future response values conditional on all covariate values, but may depend on the covariates. For such a setting, a complete-case (CC) analysis based only on subjects with

 © 2019 The International Biometric Society

fully observed data leads to a consistent estimation of the regression parameters. However, the CC analysis ignores information in the partially observed subjects and thus may not have the desired level of efficiency, especially when the proportion of subjects with missing data is not small. How to effectively use the partially observed information to improve estimation efficiency over the CC analysis is of great interest.

By modeling the missingness given both the response and the subset of fully observed covariates, Bartlett *et al*. (2014) proposed the augmented complete-case (ACC) estimator. Note that the missingness model they assumed is not for the MNAR mechanism, which depends on the subset of missing covariates as well, but is rather for the missingness probability conditional on all fully observed variables in the data set. With this model assumption, Bartlett *et al*. (2014) derived the optimal augmentation term that ensures an efficiency improvement over the CC analysis. Noting that the ACC estimating function is a simple sum of the CC analysis estimating function and an augmentation term, Xie and Zhang (2017) proposed to treat the two pieces as an over-identified estimating function and estimated the regression parameters based on the empirical likelihood method (Qin and Lawless, 1994), which essentially finds the optimal linear combination of the two pieces instead of simply summing them. Such an application of the empirical likelihood method has also been considered for MAR data. See, for example, Qin *et al*. (2009).

Both Bartlett *et al*. (2014) and Xie and Zhang (2017) assumed a model for the missingness given all the fully observed variables to improve efficiency over the CC analysis. It may be possible to model other quantities to achieve the same goal. One straightforward example is, with the observed data, to model the distribution of the response given the subset of fully observed covariates. Note that this model is different from the regression model of primary interest that models the response given all covariates. It is natural to ask how to accommodate these different model assumptions into estimation and if they are also able to extract information from the partially observed subjects. In this article, we propose a general empirical likelihood-based framework for efficiency improvement that can accommodate different model assumptions. These assumptions yield extra estimating functions in addition to the ones used for the CC analysis. We also provide some guidelines on what quantities to model for good efficiency improvement. Our suggestions have the potential to yield smaller biases compared to existing methods when the missingness probability model is incorrectly specified. We provide both arguments based on intuition and numerical results based on simulation studies. As an illustration of the proposed method, we analyze data collected from the US National Health and Nutrition Examination Survey (NHANES).

The rest of this article is organized as follows. Section 2 gives the setup and a review of relevant methods. Section 3 covers the proposed general framework. Section 4 provides some guidelines on what quantities to model to obtain better efficiency improvement. Sections 5 and 6 contain simulation studies and a data application, respectively. Some discussion is given in Section 7.

2 | SETUP AND LITERATURE REVIEW

Let Y denote the response variable and $(\boldsymbol{X}, \boldsymbol{Z})$ the vector of covariates. The model of interest is the regression of Y on $(\boldsymbol{X}, \boldsymbol{Z})$ specified by

$$E(Y|\boldsymbol{X}, \boldsymbol{Z}) = g(\boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\beta}_0), \tag{1}$$

where $g(\cdot)$ is a known monotone and continuously differentiable link function and $\boldsymbol{\beta}$ is the regression parameter with true value $\boldsymbol{\beta}_0$. When data are fully observed, a typical way of estimating $\boldsymbol{\beta}_0$ is to solve $\sum_{i=1}^{n} \boldsymbol{U}(Y_i, \boldsymbol{X}_i, \boldsymbol{Z}_i; \boldsymbol{\beta}) = \boldsymbol{0}$, where $\boldsymbol{U}(Y, \boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\beta}) = \boldsymbol{d}(\boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\beta})\epsilon(\boldsymbol{\beta})$, $\epsilon(\boldsymbol{\beta}) = Y - g(\boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\beta})$, and $\boldsymbol{d}(\boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\beta})$ is a user-specified function of $(\boldsymbol{X}, \boldsymbol{Z})$ and may depend on $\boldsymbol{\beta}$ as well. One example is

$$\boldsymbol{d}(\boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\beta}) = \frac{\partial g(\boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \mathrm{Var}(Y|\boldsymbol{X}, \boldsymbol{Z})^{-1},$$

which leads to a semiparametrically efficient estimator for $\boldsymbol{\beta}_0$ under the regression model (1) (eg, Tsiatis, 2006).

We consider the case where Y and $\boldsymbol{Z}$ are fully observed but $\boldsymbol{X}$ is subject to missingness. Let R denote an indicator variable such that $R = 1$ if $\boldsymbol{X}$ is observed and $R = 0$ if $\boldsymbol{X}$ is missing. The observed data are n independent and identically distributed copies of $(Y, R\boldsymbol{X}, \boldsymbol{Z}, R)$. In this article we consider the MNAR mechanism where the missingness of $\boldsymbol{X}$ can depend on the possibly missing $\boldsymbol{X}$ but is conditionally independent of Y given $\boldsymbol{X}$ and $\boldsymbol{Z}$; that is, $R \perp Y | (\boldsymbol{X}, \boldsymbol{Z})$. Such a MNAR mechanism is oftentimes more plausible than the MAR mechanism, especially when the response Y is measured at a later time point.

Under this MNAR mechanism, the CC analysis by solving $\sum_{i=1}^{n} R_i \boldsymbol{U}(Y_i, \boldsymbol{X}_i, \boldsymbol{Z}_i; \boldsymbol{\beta}) = \boldsymbol{0}$ yields a consistent estimator for $\boldsymbol{\beta}_0$. However, the CC estimator does not use any information from the partially observed subjects and thus may not have the desired level of estimation efficiency.

To improve efficiency over the CC analysis, additional model assumptions other than (1) need to be made. Bartlett *et al*. (2014) assumed a logistic regression model $\pi(Y, \boldsymbol{Z}; \boldsymbol{\alpha})$ for $P(R = 1|Y, \boldsymbol{Z})$, where the parameter $\boldsymbol{\alpha}$ has true value $\boldsymbol{\alpha}_0$ such that $\pi(Y, \boldsymbol{Z}; \boldsymbol{\alpha}_0) = P(R = 1|Y, \boldsymbol{Z})$. Since both Y and $\boldsymbol{Z}$ are fully observed, a consistent estimator $\hat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}_0$ can be obtained by maximizing the binomial likelihood

$$\prod_{i=1}^{n} \pi(Y_i, \boldsymbol{Z}_i; \boldsymbol{\alpha})^{R_i} \{1 - \pi(Y_i, \boldsymbol{Z}_i; \boldsymbol{\alpha})\}^{1-R_i}. \quad (2)$$

Bartlett *et al*. (2014) proposed the ACC estimator $\hat{\boldsymbol{\beta}}_{\mathrm{ACC}}$ for $\boldsymbol{\beta}_0$ by solving

$$\sum_{i=1}^{n} \{R_i \boldsymbol{U}(Y_i, \boldsymbol{X}_i, \boldsymbol{Z}_i; \boldsymbol{\beta}) + \boldsymbol{V}(Y_i, \boldsymbol{Z}_i, R_i; \boldsymbol{\beta}, \hat{\boldsymbol{\alpha}})\} = \boldsymbol{0}, \quad (3)$$

where $\boldsymbol{V}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}, \boldsymbol{\alpha}) = \{R - \pi(Y, \boldsymbol{Z}; \boldsymbol{\alpha})\}\boldsymbol{\phi}(Y, \boldsymbol{Z}; \boldsymbol{\beta})$ and $\boldsymbol{\phi}(Y, \boldsymbol{Z}; \boldsymbol{\beta})$ is a user-specified function that has the same dimension as $\boldsymbol{\beta}$. They showed that the optimal $\boldsymbol{\phi}(Y, \boldsymbol{Z}; \boldsymbol{\beta})$ that leads to the smallest asymptotic variance of $\hat{\boldsymbol{\beta}}_{\mathrm{ACC}}$ is

$$\boldsymbol{\phi}_{opt}(Y, \boldsymbol{Z}; \boldsymbol{\beta}) = -E\{\boldsymbol{U}(Y, \boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\beta})|Y, \boldsymbol{Z}, R = 1\}. \quad (4)$$

When a nonoptimal $\boldsymbol{\phi}(Y, \boldsymbol{Z}; \boldsymbol{\beta})$ is used, however, although $\hat{\boldsymbol{\beta}}_{\mathrm{ACC}}$ is still consistent, it may lose efficiency compared with the CC estimator. In this case, Bartlett *et al*. (2014) proposed a modification to (3) by considering the optimal linear combination of $R\boldsymbol{U}(Y, \boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\beta})$ and $\boldsymbol{V}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}, \hat{\boldsymbol{\alpha}})$ so that the resulting estimator, denoted by $\hat{\boldsymbol{\beta}}_{\mathrm{ACC2}}$, is at least as efficient as the CC estimator.

Noticing that both $R\boldsymbol{U}(Y, \boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\beta})$ and $\boldsymbol{V}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}, \boldsymbol{\alpha})$ in (3) have mean zero when evaluated at $\boldsymbol{\beta}_0$ and $\boldsymbol{\alpha}_0$, Xie and Zhang (2017) considered the overidentified estimating function,

$$\begin{pmatrix} R\boldsymbol{U}(Y, \boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\beta}) \\ \boldsymbol{V}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}) \end{pmatrix} \quad (5)$$

for $\boldsymbol{\beta}$. They also considered combining this estimating function with the score function for $\boldsymbol{\alpha}$ corresponding to (2) to form another over-identified estimating function

$$\begin{pmatrix} R\boldsymbol{U}(Y, \boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\beta}) \\ \boldsymbol{V}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}, \boldsymbol{\alpha}) \\ \dfrac{R - \pi(Y, Z; \boldsymbol{\alpha})}{\pi(Y, Z; \boldsymbol{\alpha})\{1 - \pi(Y, Z; \boldsymbol{\alpha})\}} \dfrac{\partial \pi(Y, Z; \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \end{pmatrix} \quad (6)$$

for $(\boldsymbol{\beta}, \boldsymbol{\alpha})$. Xie and Zhang (2017) proposed to use the empirical likelihood method (Qin and Lawless, 1994) to estimate $\boldsymbol{\beta}_0$ based on the estimating functions in (5) or (6). They showed that, when $\boldsymbol{\phi}_{opt}(Y, \boldsymbol{Z}; \boldsymbol{\beta})$ is used in $\boldsymbol{V}(Y, \boldsymbol{Z}, R, \boldsymbol{\beta}, \boldsymbol{\alpha})$, estimators based on both (5) and (6) are asymptotically equivalent to the ACC estimator. When a nonoptimal $\boldsymbol{\phi}(Y, \boldsymbol{Z}; \boldsymbol{\beta})$ is used, the estimator based on (6) is at least as efficient as both the CC estimator and the estimator based on (5), but the estimator based on (5) may be less efficient than the CC estimator. Refer to Xie and Zhang (2017) for a more detailed efficiency comparison.

3 | A GENERAL ESTIMATION FRAMEWORK

The methods in Bartlett *et al*. (2014) and Xie and Zhang (2017) represent two ways to augment the CC estimating function $R\boldsymbol{U}(Y, \boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\beta})$, and both rely on a correct model for $P(R = 1|Y, \boldsymbol{Z})$. It is possible to assume models for quantities other than $P(R = 1|Y, \boldsymbol{Z})$. We propose a general empirical likelihood-based estimation framework that can accommodate different modeling strategies.

In general, let $\boldsymbol{h}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}, \theta)$ denote a set of estimating functions for $\boldsymbol{\beta}$, which depend on the fully observed variables, Y, $\boldsymbol{Z}$, and R, and some nuisance parameter θ that is introduced when modeling quantities beyond (1). Combining $R\boldsymbol{U}(Y, \boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\beta})$ and $\boldsymbol{h}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}, \theta)$, we have an over-identified set of estimating functions for $\boldsymbol{\beta}$. Our proposed empirical likelihood-based estimator $\hat{\boldsymbol{\beta}}_{EL}$ for $\boldsymbol{\beta}_0$ is the corresponding component of the maximizer defined through

$$\max_{p_1, \ldots, p_n, \boldsymbol{\beta}, \theta} \prod_{i=1}^{n} p_i \quad \text{subject to}$$

$$p_i \geqslant 0, \quad \sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} p_i \begin{pmatrix} R_i \boldsymbol{U}(Y_i, \boldsymbol{X}_i, \boldsymbol{Z}_i; \boldsymbol{\beta}) \\ \boldsymbol{h}(Y_i, \boldsymbol{Z}_i, R_i; \boldsymbol{\beta}, \theta) \end{pmatrix} = \boldsymbol{0}. \quad (7)$$

where $p_i = dF(R_i, Y_i, \boldsymbol{X}_i, \boldsymbol{Z}_i)$, $i = 1, \ldots, n$, are a discrete distribution on the observed data. Here we require the dimension of $\boldsymbol{h}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}, \theta)$ be larger than the dimension of θ. A discussion on this point is given after Theorem 1 below.

On the basis of the results in Qin and Lawless (1994), we have the following theorem regarding the consistency and the asymptotic distribution of $\hat{\boldsymbol{\beta}}_{\mathrm{EL}}$. The derivation is given in the Supporting Information.

Theorem 1. *If* $E\{\boldsymbol{h}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}_0, \theta_0)\} = \boldsymbol{0}$ *for a unique* θ_0*, then* $\hat{\boldsymbol{\beta}}_{\mathrm{EL}}$ *is consistent and* $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{EL}} - \boldsymbol{\beta}_0)$ *has an asymptotic normal distribution with mean zero and variance*

$$\left\{E\left(R\boldsymbol{U}_{\boldsymbol{\beta}}^T\right)E(R\boldsymbol{U}\boldsymbol{U}^T)^{-1}E(R\boldsymbol{U}_{\boldsymbol{\beta}}) + \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^T\right\}^{-1} \tag{8}$$

where $\boldsymbol{U} = \boldsymbol{U}(Y, \boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\beta}_0)$, $\boldsymbol{U}_{\boldsymbol{\beta}} = \partial \boldsymbol{U}(Y, \boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\beta}_0)/\partial \boldsymbol{\beta}$, $\boldsymbol{h} = \boldsymbol{h}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}_0, \theta_0)$, $\boldsymbol{h}_{\boldsymbol{\beta}} = \partial \boldsymbol{h}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}_0, \theta_0)/\partial \boldsymbol{\beta}$, *and* $\boldsymbol{h}_{\theta} = \partial \boldsymbol{h}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}_0, \theta_0)/\partial \theta$,

$$\begin{aligned}
\boldsymbol{A} &= \left\{E\left(R\boldsymbol{U}_{\boldsymbol{\beta}}^T\right)E(R\boldsymbol{U}\boldsymbol{U}^T)^{-1}E(R\boldsymbol{U}\boldsymbol{h}^T) - E\left(\boldsymbol{h}_{\boldsymbol{\beta}}^T\right)\right\} \\
&\quad \cdot \{E(\boldsymbol{h}\boldsymbol{h}^T) - E(R\boldsymbol{h}\boldsymbol{U}^T)E(R\boldsymbol{U}\boldsymbol{U}^T)^{-1}E(R\boldsymbol{U}\boldsymbol{h}^T)\}^{-1}, \\
\boldsymbol{B} &= E(\boldsymbol{h}\boldsymbol{h}^T) - E(R\boldsymbol{h}\boldsymbol{U}^T)E(R\boldsymbol{U}\boldsymbol{U}^T)^{-1}E(R\boldsymbol{U}\boldsymbol{h}^T) - E(\boldsymbol{h}_{\theta}) \\
&\quad \times \left(E\left(\boldsymbol{h}_{\theta}^T\right)\left\{E(\boldsymbol{h}\boldsymbol{h}^T) - E(R\boldsymbol{h}\boldsymbol{U}^T)E(R\boldsymbol{U}\boldsymbol{U}^T)^{-1}E(R\boldsymbol{U}\boldsymbol{h}^T)\right\}^{-1}E(\boldsymbol{h}_{\theta})\right)^{-1} \\
&\quad \times E\left(\boldsymbol{h}_{\theta}^T\right).
\end{aligned}$$

From Lemma 1 in the Supporting Information, $\boldsymbol{B}$ is positive semidefinite and so is $\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^T$, therefore, the asymptotic variance of $\hat{\boldsymbol{\beta}}_{\mathrm{EL}}$ is no larger than that of the CC estimator, $\left\{E(R\boldsymbol{U}_{\boldsymbol{\beta}}^T)E(R\boldsymbol{U}\boldsymbol{U}^T)^{-1}E(R\boldsymbol{U}_{\boldsymbol{\beta}})\right\}^{-1}$.

It is crucial to ensure that the dimension of $\boldsymbol{h}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}, \theta)$ is larger than the dimension of θ. Only in this case does $\boldsymbol{h}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}, \theta)$ provide extra information for the estimation of $\boldsymbol{\beta}_0$ in addition to the information needed for estimating θ_0. Mathematically, if the dimension of $\boldsymbol{h}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}, \theta)$ is no larger than the dimension of θ, the constrained maximization (7) simply leads to $\hat{p}_i = 1/n$ and $\hat{\boldsymbol{\beta}}_{\mathrm{EL}}$ being the CC estimator.

It is easy to see that when assuming a correct model $\pi(Y, \boldsymbol{Z}; \boldsymbol{\alpha})$ for $P(R = 1|Y, \boldsymbol{Z})$, this general framework covers (6) as proposed in Xie and Zhang (2017), where $\boldsymbol{h}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}, \theta)$ comprises the latter two components of (6) and $\theta = \boldsymbol{\alpha}$. Note that this framework can be further extended to cover the case where part or all of θ is estimated separately and then plugged into $\boldsymbol{h}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}, \theta)$ for estimation of $\boldsymbol{\beta}$. Such an extension would cover (5) as proposed in Xie and Zhang (2017), where $\boldsymbol{h}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}, \theta) = \boldsymbol{V}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}, \boldsymbol{\alpha})$, and $\theta = \boldsymbol{\alpha}$ is estimated separately. In this article, we do not explicitly consider such an extension for two reasons. First, when θ comprises parameters from different models, there are different choices of which part is estimated separately and the asymptotic distribution of the resulting estimator for $\boldsymbol{\beta}_0$ depends on the specific choice. This makes it difficult to establish a general result for efficiency comparison. Second, estimating part of θ separately does not guarantee an efficiency improvement over the CC estimator, as shown by the results in Xie and Zhang (2017) corresponding to using (5). Refer to Section 7 for some relevant discussion.

This general framework allows the possibility of modeling quantities different from $P(R = 1|Y, \boldsymbol{Z})$ to improve efficiency over the CC analysis. A straightforward example is to model $E(Y|\boldsymbol{Z})$. For instance, assuming a model $E(Y|\boldsymbol{Z}; \boldsymbol{\gamma}) = \mu(\gamma_c + \boldsymbol{\gamma}_Z^T\boldsymbol{Z})$ with a known link function $\mu(\cdot)$ and unknown parameter $\boldsymbol{\gamma}$, we may take $\boldsymbol{h}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}, \theta)$ to be $\boldsymbol{d}(\boldsymbol{Z})\left\{Y - \mu(\gamma_c + \boldsymbol{\gamma}_Z^T\boldsymbol{Z})\right\}$ and θ to be $\boldsymbol{\gamma}$, where $\boldsymbol{d}(\boldsymbol{Z})$ is a user-specified vector function of $\boldsymbol{Z}$ with dimension larger than the dimension of $\boldsymbol{\gamma}$. When this model is correctly specified in the sense that $E(Y|\boldsymbol{Z}; \boldsymbol{\gamma}_0) = E(Y|\boldsymbol{Z})$ for $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$, Theorem 1 guarantees that $\hat{\boldsymbol{\beta}}_{\mathrm{EL}}$ is more efficient than the CC estimator. Another example is to model both $P(R = 1|Y, \boldsymbol{Z})$ and $E(Y|\boldsymbol{Z})$. In this case, we could take $\boldsymbol{h}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}, \theta)$ to be

$$\begin{pmatrix} \boldsymbol{V}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}, \boldsymbol{\alpha}) \\ \dfrac{R - \pi(\boldsymbol{\alpha})}{\pi(\boldsymbol{\alpha})\{1 - \pi(\boldsymbol{\alpha})\}}\dfrac{\partial \pi(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \\ \boldsymbol{d}(\boldsymbol{Z})\left\{Y - \mu\left(\gamma_c + \boldsymbol{\gamma}_Z^T\boldsymbol{Z}\right)\right\} \end{pmatrix}$$

and θ to be $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$. Consistency and efficiency improvement over the CC analysis, in this case, requires both $P(R = 1|Y, \boldsymbol{Z})$ and $E(Y|\boldsymbol{Z})$ to be correctly modeled.

Model compatibility issues may arise when modeling additional quantities since we have already assumed a model of interest (1). For example, (1) may impose some restrictions on how to model $E(Y|\boldsymbol{Z})$. When (1) is a linear model with $g(\cdot)$ the identity link function, $E(Y|\boldsymbol{Z})$ may also be taken as a linear model with $\mu(\cdot)$ the identity link. When $g(\cdot)$ is the logit link; however, taking $\mu(\cdot)$ to be the logit link usually does not lead to a model for $E(Y|\boldsymbol{Z})$ that is mathematically compatible with (1), and it is difficult or impossible to find a link function $\mu(\cdot)$ that leads to mathematical compatibility. Our attitude on this issue is that, because most parametric models are not totally "correct" in the real world, there is always some degree of misspecification, even when the models are mathematically compatible. In practice, we use "working" models to fit the data. These models should not deviate substantially from the observed data, and we support this through model checking and model diagnosis. Model compatibility thus has to be taken with a grain of salt. Therefore, our attitude is to seek models that are mathematically compatible if possible, but in any case, ones that are sufficiently consistent with the observed data. Model-checking and diagnosis techniques can be used to reduce the chance of serious model incompatibility.

4 | CHOICES OF QUANTITIES TO MODEL

The efficiency improvement over the CC analysis implied by Theorem 1 is achieved by making model assumptions in addition to the model of interest in (1). Bartlett *et al.* (2014) and Xie and Zhang (2017) assumed a model for $P(R=1|Y,\boldsymbol{Z})$. Other model assumptions can be considered as well. Different assumptions involve different amounts of information and thus lead to different efficiency improvements over the CC analysis. Although it is natural to ask what quantities should be modeled in order to have the most improvement, providing an answer is tremendously challenging, if not impossible, since even for the two cases of modeling $P(R=1|Y,\boldsymbol{Z})$ and $E(Y|\boldsymbol{Z})$ there does not seem to be a direct efficiency comparison. Note, for example, the complex dependence of $\boldsymbol{ABA}^T$ in (8) on $\boldsymbol{\beta}$ and $\boldsymbol{h}(Y,\boldsymbol{Z},R;\boldsymbol{\beta},\theta)$. Such a generally non-simplifiable dependence makes it almost impossible to find the "best" quantity to model. Compounding the problem is the incompatibility issue wherein many settings there is no mathematically compatible model. In this case, models that agree closely with the observed data will presumably lead to estimates with small bias and efficiency improvement, but this needs to be investigated using numerical studies.

To gain insight into what quantities should be modeled we consider a simpler situation by dropping the dependence of $\boldsymbol{h}(Y,\boldsymbol{Z},R;\boldsymbol{\beta},\theta)$ on R, $\boldsymbol{\beta}$, and θ. In other words, we find the optimal estimating function $\boldsymbol{h}(Y,\boldsymbol{Z})$ with $E\{\boldsymbol{h}(Y,\boldsymbol{Z})\}=\boldsymbol{0}$ under the true underlying distribution. From Theorem 1, with $\boldsymbol{h}(Y,\boldsymbol{Z},R;\boldsymbol{\beta},\theta)$ replaced by $\boldsymbol{h}(Y,\boldsymbol{Z})$, the asymptotic variance in (8) becomes $\left[E(R\boldsymbol{U}_\beta^T)\{\mathrm{Var}(\mathrm{Resid}(R\boldsymbol{U},\boldsymbol{h}))\}^{-1}E(R\boldsymbol{U}_\beta)\right]^{-1}$, where $\mathrm{Resid}(R\boldsymbol{U},\boldsymbol{h})=R\boldsymbol{U}-E(R\boldsymbol{U}\boldsymbol{h}^T)E(\boldsymbol{h}\boldsymbol{h}^T)^{-1}\boldsymbol{h}$ is the residual of the projection of $R\boldsymbol{U}$ on the linear space spanned by $\boldsymbol{h}$. Due to this special structure, simple algebra shows that the optimal $\boldsymbol{h}(Y,\boldsymbol{Z})$ leading to the most efficiency improvement over the CC analysis is given by

$$\begin{aligned}\boldsymbol{h}_{\mathrm{opt}}(Y,\boldsymbol{Z}) &= E\{R\boldsymbol{U}(Y,\boldsymbol{X},\boldsymbol{Z};\boldsymbol{\beta}_0)|Y,\boldsymbol{Z}\}\\ &= P(R=1|Y,\boldsymbol{Z})\\ &\quad\times E\{\boldsymbol{U}(Y,\boldsymbol{X},\boldsymbol{Z};\boldsymbol{\beta}_0)|Y,\boldsymbol{Z},R=1\}.\end{aligned}$$

However, $\boldsymbol{h}_{\mathrm{opt}}(Y,\boldsymbol{Z})$ is not directly applicable due to its dependence on the unknown underlying data distribution. First, it depends on the data distribution through the unknown $\boldsymbol{\beta}_0$. To overcome this, we consider the estimating function $P(R=1|Y,\boldsymbol{Z})E\{\boldsymbol{U}(Y,\boldsymbol{X},\boldsymbol{Z};\boldsymbol{\beta})|Y,\boldsymbol{Z},R=1\}$ instead of $\boldsymbol{h}_{\mathrm{opt}}(Y,\boldsymbol{Z})$. Second, $\boldsymbol{h}_{\mathrm{opt}}(Y,\boldsymbol{Z})$ depends on the data distribution through the unknown $P(R=1|Y,\boldsymbol{Z})$ and $f(\boldsymbol{X}|Y,\boldsymbol{Z},R=1)$. To overcome this, we assume models $\pi(Y,\boldsymbol{Z};\boldsymbol{\alpha})=P(R=1|Y,\boldsymbol{Z};\boldsymbol{\alpha})$ and $f(\boldsymbol{X}|Y,\boldsymbol{Z},R=1;\boldsymbol{\gamma})$ that depend on nuisance parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$. Based on these considerations, the auxiliary estimating function we suggest is

$$\begin{aligned}&\boldsymbol{h}_{\mathrm{use}}(Y,\boldsymbol{Z};\boldsymbol{\beta},\theta)\\ &\quad=\pi(Y,\boldsymbol{Z};\boldsymbol{\alpha})E\{\boldsymbol{U}(Y,\boldsymbol{X},\boldsymbol{Z};\boldsymbol{\beta})|Y,\boldsymbol{Z},R=1;\boldsymbol{\gamma}\},\end{aligned}$$

where $\theta=(\boldsymbol{\alpha},\boldsymbol{\gamma})$ and $E\{\boldsymbol{U}(Y,\boldsymbol{X},\boldsymbol{Z};\boldsymbol{\beta})|Y,\boldsymbol{Z},R=1;\boldsymbol{\gamma}\}$ is taken under the model $f(\boldsymbol{X}|Y,\boldsymbol{Z},R=1;\boldsymbol{\gamma})$. It is easy to verify that $E\{\boldsymbol{h}_{\mathrm{use}}(Y,\boldsymbol{Z};\boldsymbol{\beta}_0,\theta_0)\}=\boldsymbol{0}$, where $\theta_0=(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0)$ and $\boldsymbol{\gamma}_0$ is the true value of $\boldsymbol{\gamma}$ such that $f(\boldsymbol{X}|Y,\boldsymbol{Z},R=1;\boldsymbol{\gamma}_0)=f(\boldsymbol{X}|Y,\boldsymbol{Z},R=1)$. Based on reasons given below Theorem 1, we consider estimating $\boldsymbol{\alpha}_0$ and $\boldsymbol{\gamma}_0$ jointly with $\boldsymbol{\beta}_0$. This consideration leads to replacing $\boldsymbol{h}(Y,\boldsymbol{Z},R;\boldsymbol{\beta},\theta)$ in (7) by

$$\begin{pmatrix}\boldsymbol{h}_{\mathrm{use}}(Y,\boldsymbol{Z};\boldsymbol{\beta},\theta)\\ \dfrac{R-\pi(Y,Z;\boldsymbol{\alpha})}{\pi(Y,Z;\boldsymbol{\alpha})\{1-\pi(Y,Z;\boldsymbol{\alpha})\}}\dfrac{\partial\pi(Y,Z;\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\\ R\boldsymbol{S}(Y,\boldsymbol{X},\boldsymbol{Z};\boldsymbol{\gamma})\end{pmatrix}, \tag{9}$$

where the second component is the score function corresponding to (2) for estimating $\boldsymbol{\alpha}_0$ and $\boldsymbol{S}(Y,\boldsymbol{X},\boldsymbol{Z};\boldsymbol{\gamma})$ is a user-specified estimating function for estimating $\boldsymbol{\gamma}_0$ such that $E\{R\boldsymbol{S}(Y,\boldsymbol{X},\boldsymbol{Z};\boldsymbol{\gamma}_0)\}=\boldsymbol{0}$. For example, $\boldsymbol{S}(Y,\boldsymbol{X},\boldsymbol{Z};\boldsymbol{\gamma})$ may be taken to be the score function corresponding to the model $f(\boldsymbol{X}|Y,\boldsymbol{Z},R=1;\boldsymbol{\gamma})$.

Implementation based on (9) involves two model assumptions in addition to (1), one for $P(R=1|Y,\boldsymbol{Z})$ and one for $f(\boldsymbol{X}|Y,\boldsymbol{Z},R=1)$. Both models need to be correctly specified for the proposed estimator $\hat{\boldsymbol{\beta}}_{EL}$ to be consistent. In comparison, the ACC estimator in Bartlett *et al.* (2014) treats the model $f(\boldsymbol{X}|Y,\boldsymbol{Z},R=1;\boldsymbol{\gamma})$ as a working model and its consistency only requires correct specification of $\pi(Y,\boldsymbol{Z};\boldsymbol{\alpha})$. However, when $f(\boldsymbol{X}|Y,\boldsymbol{Z},R=1;\boldsymbol{\gamma})$ is incorrectly specified, the ACC estimator may be less efficient than the CC estimator. Since the main objective is to improve efficiency over the CC estimator because it is already consistent, $f(\boldsymbol{X}|Y,\boldsymbol{Z},R=1;\boldsymbol{\gamma})$ still needs to be a "good" model for the ACC method, if not the "correct" one. In contrast, as discussed at the end of Section 3, in the real world there is always some degree of misspecification for parametric models. Therefore, we think that (9) is also worth consideration in scenarios where the ACC method is expected to provide an improvement over the CC analysis. Note that the model for $f(\boldsymbol{X}|Y,\boldsymbol{Z},R=1)$ is fitted based on the complete cases. Complications for specifying, fitting, and checking this model may arise when $\boldsymbol{X}$ is multivariate, especially if it is a mix of continuous and discrete variables.

When the dimension of $\boldsymbol{\beta}$ is larger than that of $\boldsymbol{\gamma}$, $R\boldsymbol{S}(Y, \boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\gamma})$ in (9) may be dropped in the implementation, because in this case $R\boldsymbol{U}(Y, \boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\beta})$ combined with the first two components of (9) already provides a set of over-identified estimating functions for $(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{\gamma}_0)$. The benefit of dropping $R\boldsymbol{S}(Y, \boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\gamma})$ from (9) in this case is twofold. First, the reduction of the total number of estimating functions may improve the numerical performance of the empirical likelihood method, especially when this number is large. Second and more importantly, it will substantially reduce the bias of $\hat{\boldsymbol{\beta}}_{\mathrm{EL}}$ when $f(\boldsymbol{X}|Y, \boldsymbol{Z}, R = 1; \boldsymbol{\gamma})$ is misspecified. The reason is that, when $f(\boldsymbol{X}|Y, \boldsymbol{Z}, R = 1; \boldsymbol{\gamma})$ is misspecified, $R\boldsymbol{S}(Y, \boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\gamma})$ provides "incorrect" information about the data distribution. When this "incorrect" information is accommodated in calculating $\hat{\boldsymbol{\beta}}_{\mathrm{EL}}$ and $\hat{\boldsymbol{\gamma}}$, it pulls $\hat{\boldsymbol{\beta}}_{\mathrm{EL}}$ away from the true value $\boldsymbol{\beta}_0$. Dropping $R\boldsymbol{S}(Y, \boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\gamma})$ removes this undesired impact. On the contrary, the ACC and Xie and Zhang's (2017) methods still require $R\boldsymbol{S}(Y, \boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\gamma})$ as the estimating function to estimate $\boldsymbol{\gamma}$, and thus still make full use of this "incorrect" information. Because of this, our proposed estimator can become less biased than the ACC and Xie and Zhang's (2017) when the model for $P(R = 1|Y, \boldsymbol{Z})$ is also misspecified. Simulation Study 2 in Section 5 provides numerical evidence supporting this intuition. This observation is of high importance because, in the real world, it is likely that models for $P(R = 1|Y, \boldsymbol{Z})$ and $f(\boldsymbol{X}|Y, \boldsymbol{Z}, R = 1)$ are both misspecified and none of the existing estimators is consistent. Thus a possibly smaller bias by our proposed method becomes highly desired.

We also note that $\boldsymbol{h}_{use}(Y, \boldsymbol{Z}; \boldsymbol{\beta}, \theta)$ does not have a rigorous theoretical justification, and (9) is not necessarily the "optimal" estimating function in theory. Although using (9) is guaranteed to improve efficiency over the CC analysis when corresponding models are correctly specified, there is not a direct efficiency comparison to the ACC method.

5 | SIMULATION STUDIES

5.1 | Study 1

This simulation study uses the setup in Bartlett *et al.* (2014). The data are generated as $R \sim \mathrm{Bernoulli}(0.5)$ and

$$\begin{pmatrix} Y \\ X \\ Z \end{pmatrix} \Bigg| \; R \sim \mathcal{N}\left(\begin{pmatrix} 0.2R \\ R \\ 0 \end{pmatrix}, \; \begin{pmatrix} 1 & 0.25 & 0.25 \\ 0.25 & 1 & 0.25 \\ 0.25 & 0.25 & 1 \end{pmatrix} \right),$$

and the observed data vector is (Y, RX, Z, R). This data generating process implies that the missingness of X is MNAR and $R \perp Y|(X, Z)$. In addition, it ensures that $P(R = 1|Y, Z)$ can be correctly modeled by a logistic regression. The conditional mean model of interest is $E(Y|X, Z) = \beta_c + \beta_X X + \beta_Z Z$ with $\boldsymbol{\beta}_0 = (\beta_c, \beta_X, \beta_Z) = (0, 0.2, 0.2)$. This simulation takes

$$\boldsymbol{U}(Y, X, Z; \boldsymbol{\beta}) = (1, X, Z)^{\mathrm{T}}(Y - \beta_c - \beta_X X - \beta_Z Z).$$

Following Bartlett *et al.* (2014), let $\mathrm{logit}\{\pi(Y, Z; \boldsymbol{\alpha})\} = \alpha_c + \alpha_Y Y + \alpha_Z Z$ be the correctly specified model for $P(R = 1|Y, Z)$, $f_1(X|Y, Z, R = 1; \boldsymbol{\gamma})$ the correctly specified model $\mathcal{N}(\gamma_c + \gamma_Y Y + \gamma_Z Z, \gamma_\sigma^2)$ for $f(X|Y, Z, R = 1)$, and $f_2(X|Y, Z, R = 1; \boldsymbol{\gamma})$ the misspecified model $\mathcal{N}(\gamma_c + \gamma_Y Y^2 + \gamma_Z Z^2, \gamma_\sigma^2)$ for $f(X|Y, Z, R = 1)$. The two models for $f(X|Y, Z, R = 1)$ are used to calculate $\boldsymbol{\phi}_{\mathrm{opt}}(Y, Z; \boldsymbol{\beta})$ in (4).

We present the performance of the following estimators.

1. The CC analysis estimator $\hat{\boldsymbol{\beta}}_{\mathrm{CC}}$.
2. Two ACC estimators $\hat{\boldsymbol{\beta}}_{\mathrm{ACC}-1}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{ACC}-2}$, both of which use $\pi(Y, Z; \boldsymbol{\alpha})$, but $\hat{\boldsymbol{\beta}}_{\mathrm{ACC}-1}$ is based on $f_1(X|Y, Z, R = 1; \boldsymbol{\gamma})$ and $\hat{\boldsymbol{\beta}}_{\mathrm{ACC}-2}$ is based on $f_2(X|Y, Z, R = 1; \boldsymbol{\gamma})$.
3. Two ACC2 estimators $\hat{\boldsymbol{\beta}}_{\mathrm{ACC2}-1}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{ACC2}-2}$ as proposed in Bartlett *et al.* (2014), based on the ACC estimators $\hat{\boldsymbol{\beta}}_{\mathrm{ACC}-1}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{ACC}-2}$, respectively.
4. Two estimators from Xie and Zhang (2017) $\hat{\boldsymbol{\beta}}_{\mathrm{XZ1}-1}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{XZ1}-2}$ based on (5), using the same models as those for $\hat{\boldsymbol{\beta}}_{\mathrm{ACC}-1}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{ACC}-2}$, respectively.
5. Two estimators from Xie and Zhang (2017) $\hat{\boldsymbol{\beta}}_{\mathrm{XZ2}-1}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{XZ2}-2}$ based on (6), using the same models as those for $\hat{\boldsymbol{\beta}}_{\mathrm{ACC}-1}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{ACC}-2}$, respectively.
6. Two estimators $\hat{\boldsymbol{\beta}}_{\mathrm{EL}-1}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{EL}-2}$ based on our proposed method with (9), using the same models as those for $\hat{\boldsymbol{\beta}}_{\mathrm{ACC}-1}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{ACC}-2}$, respectively.

For $\hat{\boldsymbol{\beta}}_{\mathrm{EL}-1}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{EL}-2}$, the $\boldsymbol{S}(Y, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\gamma})$ in (9) is taken to be

$$\begin{pmatrix} (1, Y, Z)^{\mathrm{T}}(X - \gamma_c - \gamma_Y Y - \gamma_Z Z) \\ (X - \gamma_c - \gamma_Y Y - \gamma_Z Z)^2 - \gamma_\sigma^2 \end{pmatrix}$$

and

$$\begin{pmatrix} (1, Y^2, Z^2)^{\mathrm{T}}(X - \gamma_c - \gamma_Y Y^2 - \gamma_Z Z^2) \\ (X - \gamma_c - \gamma_Y Y^2 - \gamma_Z Z^2)^2 - \gamma_\sigma^2 \end{pmatrix},$$

respectively.

Table 1 summarizes the simulation results based on 1000 replications. It is seen that, EL-1 based on correctly specified models performs equally well compared to the ACC and Xie and Zhang's (2017) estimators using the same models, and all have improved efficiency over the CC

TABLE 1 Simulation results for study 1

	Bias (empirical standard error) [RMSE]		
	β_c ($\beta_{c0} = 0$)	β_X ($\beta_{X0} = 0.2$)	β_Z ($\beta_{Z0} = 0.2$)
$n = 400$			
CC	0.001 (0.094) [0.094]	0.001 (0.071) [0.071]	0.001 (0.070) [0.070]
ACC-1	0.001 (0.093) [0.093]	0.002 (0.069) [0.069]	0.001 (0.052) [0.052]
ACC-2	0.003 (0.094) [0.094]	−0.001 (0.072) [0.072]	0.002 (0.053) [0.053]
ACC2-1	0.005 (0.093) [0.093]	−0.001 (0.069) [0.069]	0.001 (0.053) [0.053]
ACC2-2	0.007 (0.095) [0.095]	−0.002 (0.070) [0.070]	0.000 (0.054) [0.054]
XZ1-1	0.003 (0.093) [0.093]	−0.001 (0.069) [0.069]	0.001 (0.053) [0.053]
XZ1-2	0.001 (0.095) [0.095]	0.000 (0.072) [0.072]	0.001 (0.054) [0.054]
XZ2-1	0.003 (0.093) [0.093]	−0.001 (0.069) [0.069]	0.001 (0.053) [0.053]
XZ2-2	0.002 (0.095) [0.095]	0.000 (0.072) [0.072]	0.001 (0.054) [0.054]
EL-1	0.001 (0.094) [0.094]	0.001 (0.070) [0.070]	0.001 (0.053) [0.053]
EL-2	0.158 (0.072) [0.173]	−0.151 (0.067) [0.165]	0.045 (0.052) [0.068]
$n = 1000$			
CC	0.002 (0.064) [0.064]	−0.001 (0.045) [0.045]	−0.001 (0.043) [0.043]
ACC-1	0.001 (0.063) [0.063]	0.000 (0.045) [0.045]	−0.001 (0.032) [0.032]
ACC-2	0.002 (0.065) [0.065]	−0.001 (0.046) [0.046]	0.000 (0.032) [0.032]
ACC2-1	0.003 (0.064) [0.064]	−0.001 (0.045) [0.045]	−0.001 (0.032) [0.032]
ACC2-2	0.004 (0.064) [0.064]	−0.002 (0.045) [0.045]	−0.001 (0.033) [0.033]
XZ1-1	0.002 (0.063) [0.063]	−0.002 (0.045) [0.045]	−0.001 (0.032) [0.032]
XZ1-2	0.001 (0.064) [0.064]	−0.001 (0.045) [0.045]	−0.001 (0.032) [0.032]
XZ2-1	0.003 (0.063) [0.063]	−0.002 (0.045) [0.045]	−0.001 (0.032) [0.032]
XZ2-2	0.002 (0.064) [0.064]	−0.001 (0.045) [0.045]	−0.001 (0.032) [0.032]
EL-1	0.001 (0.064) [0.064]	0.000 (0.045) [0.045]	−0.001 (0.032) [0.032]
EL-2	0.155 (0.166) [0.059]	−0.151 (0.051) [0.159]	0.043 (0.033) [0.054]

estimator. Note that in this case the ACC estimating equation in (3) represents the best linear combination of $R\boldsymbol{U}(Y, \boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\beta})$ and $\boldsymbol{V}(Y, \boldsymbol{Z}, R; \boldsymbol{\beta}, \boldsymbol{\alpha})$, and thus the corresponding ACC estimator has the maximum efficiency. It is also seen that EL-2 based on the misspecified model $f_2(X|Y, Z, R = 1; \boldsymbol{\gamma})$ is biased. However, we would like to point out that $f_2(X|Y, Z, R = 1; \boldsymbol{\gamma})$ is unlikely to be chosen as a model for $f(X|Y, Z, R = 1)$ in the real world. It includes quadratic effects of Y and Z without any linear effects. The likelihood ratio test comparing models $f_1(X|Y, Z, R = 1; \boldsymbol{\gamma})$ and $f_2(X|Y, Z, R = 1; \boldsymbol{\gamma})$ to the normal linear regression with Y, Z, YZ, Y^2, and Z^2 as regressors rejected the two models 60 and 985 times out of 1000 replications when $n = 400$, respectively, and these numbers became 52 and 1000 when $n = 1000$, showing that it would be extremely unlikely to choose $f_2(X|Y, Z, R = 1; \boldsymbol{\gamma})$ to model $f(X|Y, Z, R = 1)$. Therefore the bias of EL-2 in this scenario should not be interpreted exclusively as a sign against our proposed method but rather an indication of the need for a model consistent with the observed data.

5.2 | Study 2

This study considers three covariates, $X \sim \text{Exponential}(2)$, $W \sim N(0, 1)$, and $Z|W \sim N(W, 1)$. Given the covariates, Y is generated as $Y = \beta_c + \beta_X X + \beta_Z Z + \beta_W W + \epsilon$, where $\boldsymbol{\beta}_0 = (\beta_c, \beta_X, \beta_Z, \beta_W) = (0, 1, 1, 1)$ and $\epsilon \sim N(0, 1)$ is independent of the covariates X, W, and Z. The missingness of X is generated as $P(R = 1|Y, X, Z, W) = \text{expit}(1 - 0.5X + Z + W)$, under which about 50% of subjects have missing X. The conditional mean model of interest is $E(Y|X, Z, W) = \beta_c + \beta_X X + \beta_Z Z + \beta_W W$, and this simulation takes

$$\boldsymbol{U}(Y, X, Z, W; \boldsymbol{\beta}) = (1, X, Z, W)^{\mathrm{T}}(Y - \beta_c - \beta_X X - \beta_Z Z - \beta_W W),$$

In this simulation setting, it is very challenging, if not impossible, to derive a correct model for $P(R = 1|Y, Z, W)$. We consider the logistic regression model

$$\text{logit}\{\pi(Y, Z, W; \boldsymbol{\alpha})\} = \alpha_c + \alpha_Y Y + \alpha_Z Z + \alpha_W W,$$

which is misspecified. To assess the goodness-of-fit of this model (Model 1) to the observed data, we compare it to two more complex models; one is a logistic regression with all the main effects and two way interactions of Y, Z, and W (Model 2), and the other is the generalized additive model (Hastie and Tibshirani, 1990) with logit link and all main effects of Y, Z, and W smoothed by fourth-order splines (Model 3). Taking $n = 400$ out of 1000 replications, the likelihood ratio test rejected Model 1 54 times when comparing it to Model 2 and 143 times when comparing it to Model 3, and the numbers of rejections became 44 and 136 with $n = 1000$. Therefore, model $\pi(Y, Z, W; \boldsymbol{\alpha})$ would not be rejected most of the time.

For $f(X|Y, Z, W, R = 1)$, it is also very difficult to specify the correct model. Instead, we consider the following three models: $f_1(X|Y, Z, W, R = 1; \boldsymbol{\gamma})$ is the truncated normal distribution $N(\gamma_c + \gamma_Y Y + \gamma_Z Z + \gamma_W W, \gamma_\sigma^2) I(X > 0)$, $f_2(X|Y, Z, W, R = 1; \boldsymbol{\gamma})$ is the truncated normal distribution $N(\gamma_c + \gamma_Y Y, \gamma_\sigma^2) I(X > 0)$ and $f_3(X|Y, Z, W, R = 1; \boldsymbol{\gamma})$ is the normal distribution $N(\gamma_c + \gamma_Y Y, \gamma_\sigma^2)$. These models are used to calculate $\boldsymbol{\phi}_{opt}(Y, Z, W; \boldsymbol{\beta})$ in (4). Figure 1 shows a typical P-P plot of these three models based on one simulation with $n = 400$. It clearly indicates that $f_3(X|Y, Z, W, R = 1; \boldsymbol{\gamma})$ is not a good model and is inferior to the other two. Samples with $n = 1000$ yield similar plots.

To further assess the goodness-of-fit of $f_1(X|Y, Z, W, R = 1; \boldsymbol{\gamma})$ and $f_2(X|Y, Z, W, R = 1; \boldsymbol{\gamma})$, we compare them to two more complex models; one is normal linear regression left truncated at 0 with all the main effects and two way interactions of Y, Z, and W (Model 4), and the other is normal linear regression left truncated at 0 with all the main and quadratic effects of Y, Z, and W (Model 5). Taking $n = 400$ out of 1000 replications, the likelihood ratio test rejected $f_1(X|Y, Z, W, R = 1; \boldsymbol{\gamma})$ five times and $f_2(X|Y, Z, W, R = 1; \boldsymbol{\gamma})$ 406 times when compared to Model 4 and 7 and 399 times when compared to Model 5. These numbers became 2, 981, 7, and 980 with $n = 1000$. In addition, the likelihood ratio test comparing $f_2(X|Y, Z, W, R = 1; \boldsymbol{\gamma})$ with $f_1(X|Y, Z, W, R = 1; \boldsymbol{\gamma})$ rejected the former 567 times with $n = 400$ and 994 times with $n = 1000$ out of 1000 replications. These tests suggest that $f_1(X|Y, Z, W, R = 1; \boldsymbol{\gamma})$ seems an adequate model whereas $f_2(X|Y, Z, W, R = 1; \boldsymbol{\gamma})$ is not.

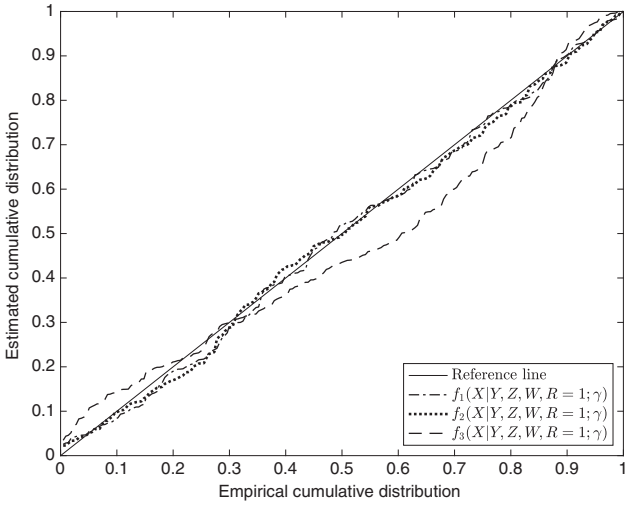

FIGURE 1 P-P plot for the three models for $f(X|Y, Z, W, R = 1)$ with $n = 400$

Table 2 summarizes the simulation results based on 1000 replications. The CC, ACC, ACC2, XZ1, XZ2, and EL estimators follow the same notation used in Study 1, now with three models $f_1(X|Y, Z, W, R = 1; \boldsymbol{\gamma})$, $f_2(X|Y, Z, W, R = 1; \boldsymbol{\gamma})$, and $f_3(X|Y, Z, W, R = 1; \boldsymbol{\gamma})$ considered. The $\boldsymbol{S}(Y, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\gamma})$ for estimating $\boldsymbol{\gamma}$ for all estimators is taken to be the score functions for these three models. Noting that $f_2(X|Y, Z, W, R = 1; \boldsymbol{\gamma})$ and $f_3(X|Y, Z, W, R = 1; \boldsymbol{\gamma})$ are clearly inadequate based on our model checking and $\boldsymbol{\gamma}$ has lower dimension than $\boldsymbol{\beta}$ for these two models, the two estimators EL2-2 and EL2-3 drop the $\boldsymbol{S}(Y, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\gamma})$ in (9), as discussed in Section 4. It is seen that, as expected, the ACC, ACC2, XZ1, XZ2, and EL estimators are all biased since neither $P(R = 1|Y, Z, W)$ nor $f(X|Y, Z, W, R = 1)$ is correctly modeled by any of the models under consideration, albeit the levels of bias vary somewhat. The ACC-3 and ACC2-3 estimators based on a clearly inadequate model $f_3(X|Y, Z, W, R = 1; \boldsymbol{\gamma})$ surprisingly have smaller bias than ACC-1 and ACC2-1 estimators based on a better model $f_1(X|Y, Z, W, R = 1; \boldsymbol{\gamma})$, but this might be just a numerical coincidence under this simulation setting. In addition, estimators EL2-2 and EL2-3 have very small bias, providing some numerical evidence supporting our intuition in Section 4. We have done some further numerical studies with different sets of parameter values that allow variation in the strength of the X effect in both the conditional mean model of interest and the missingness mechanism, and they all perform very similarly to the results in Table 2. We would like to point out, however, that the small bias of EL2-2 and EL2-3 does not have a rigorous theoretical justification and this simulation study only covers one set of scenarios. Further empirical studies motivated by real settings are recommended.

6 | DATA APPLICATION

As an application, we analyze the data collected in the year 2003 to 2004 from the NHANES. NHANES is a program conducted by the Centers for Disease Control

TABLE 2 Simulation results for study 2

	Bias (empirical standard error) [RMSE]			
	β_c ($\beta_{c0} = 0$)	β_X ($\beta_{X0} = 1$)	β_Z ($\beta_{Z0} = 1$)	β_W ($\beta_{W0} = 1$)
$n = 400$				
CC	0.004 (0.222) [0.222]	−0.007 (0.101) [0.101]	0.006 (0.149) [0.149]	−0.001 (0.206) [0.206]
ACC-1	−0.168 (0.208) [0.267]	0.060 (0.096) [0.113]	0.036 (0.139) [0.144]	0.029 (0.190) [0.192]
ACC-2	−0.161 (0.213) [0.267]	0.059 (0.099) [0.116]	0.036 (0.140) [0.145]	0.030 (0.191) [0.193]
ACC-3	0.065 (0.243) [0.252]	0.003 (0.100) [0.100]	−0.050 (0.213) [0.219]	−0.058 (0.302) [0.308]
ACC2-1	−0.142 (0.213) [0.257]	0.019 (0.101) [0.103]	0.044 (0.143) [0.150]	0.036 (0.193) [0.196]
ACC2-2	−0.135 (0.215) [0.254]	0.015 (0.101) [0.102]	0.042 (0.143) [0.149]	0.034 (0.193) [0.196]
ACC2-3	−0.043 (0.218) [0.222]	−0.005 (0.101) [0.101]	0.030 (0.142) [0.145]	0.024 (0.192) [0.194]
XZ1-1	−0.225 (0.226) [0.319]	0.105 (0.110) [0.152]	0.036 (0.145) [0.149]	0.029 (0.198) [0.200]
XZ1-2	−0.189 (0.217) [0.288]	0.088 (0.105) [0.137]	0.032 (0.142) [0.146]	0.026 (0.195) [0.197]
XZ1-3	−0.150 (0.207) [0.256]	0.063 (0.099) [0.118]	0.029 (0.141) [0.144]	0.024 (0.193) [0.194]
XZ2-1	−0.162 (0.232) [0.283]	0.043 (0.108) [0.116]	0.042 (0.146) [0.152]	0.036 (0.199) [0.202]
XZ2-2	−0.170 (0.232) [0.288]	0.048 (0.107) [0.117]	0.042 (0.145) [0.151]	0.033 (0.198) [0.200]
XZ2-3	−0.204 (0.217) [0.298]	0.061 (0.101) [0.119]	0.045 (0.145) [0.152]	0.040 (0.196) [0.200]
EL-1	−0.177 (0.226) [0.287]	0.048 (0.108) [0.118]	0.048 (0.145) [0.153]	0.042 (0.198) [0.202]
EL-2	0.310 (0.219) [0.379]	−0.153 (0.113) [0.190]	−0.048 (0.145) [0.153]	−0.055 (0.195) [0.202]
EL-3	0.234 (0.201) [0.309]	−0.090 (0.096) [0.132]	−0.059 (0.134) [0.146]	−0.063 (0.185) [0.195]
EL2-2	0.012 (0.223) [0.223]	−0.013 (0.101) [0.102]	0.005 (0.134) [0.134]	0.002 (0.188) [0.188]
EL2-3	0.011 (0.223) [0.223]	−0.013 (0.101) [0.102]	0.006 (0.135) [0.135]	0.001 (0.189) [0.189]
$n = 1000$				
CC	−0.003 (0.136) [0.136]	0.002 (0.061) [0.061]	−0.001 (0.094) [0.094]	−0.001 (0.130) [0.130]
ACC-1	−0.170 (0.127) [0.212]	0.064 (0.059) [0.087]	0.030 (0.087) [0.092]	0.028 (0.119) [0.123]
ACC-2	−0.166 (0.129) [0.210]	0.065 (0.060) [0.088]	0.032 (0.089) [0.094]	0.029 (0.120) [0.124]
ACC-3	0.056 (0.149) [0.159]	0.012 (0.062) [0.063]	−0.058 (0.133) [0.145]	−0.055 (0.190) [0.197]
ACC2-1	−0.150 (0.130) [0.199]	0.036 (0.061) [0.071]	0.034 (0.089) [0.095]	0.033 (0.121) [0.126]
ACC2-2	−0.141 (0.131) [0.193]	0.032 (0.062) [0.070]	0.033 (0.089) [0.094]	0.031 (0.120) [0.124]
ACC2-3	−0.050 (0.136) [0.145]	0.006 (0.062) [0.062]	0.024 (0.090) [0.093]	0.020 (0.124) [0.125]
XZ1-1	−0.210 (0.134) [0.249]	0.095 (0.065) [0.115]	0.030 (0.091) [0.095]	0.030 (0.123) [0.126]
XZ1-2	−0.189 (0.132) [0.230]	0.088 (0.064) [0.109]	0.027 (0.090) [0.094]	0.026 (0.121) [0.124]
XZ1-3	−0.156 (0.127) [0.201]	0.067 (0.061) [0.091]	0.025 (0.090) [0.093]	0.023 (0.121) [0.123]
XZ2-1	−0.149 (0.142) [0.206]	0.042 (0.071) [0.082]	0.035 (0.090) [0.097]	0.033 (0.122) [0.127]
XZ2-2	−0.169 (0.143) [0.222]	0.058 (0.068) [0.090]	0.033 (0.090) [0.096]	0.031 (0.121) [0.125]
XZ2-3	−0.198 (0.132) [0.238]	0.070 (0.063) [0.094]	0.037 (0.091) [0.099]	0.034 (0.123) [0.128]
EL-1	−0.181 (0.136) [0.226]	0.062 (0.065) [0.090]	0.038 (0.092) [0.100]	0.037 (0.124) [0.129]
EL-2	0.295 (0.158) [0.335]	−0.151 (0.081) [0.172]	−0.046 (0.094) [0.105]	−0.051 (0.129) [0.139]
EL-3	0.211 (0.128) [0.247]	−0.074 (0.061) [0.096]	−0.059 (0.083) [0.102]	−0.064 (0.118) [0.134]
EL2-2	0.008 (0.145) [0.145]	−0.005 (0.068) [0.068]	0.000 (0.086) [0.086]	−0.004 (0.117) [0.118]
EL2-3	0.005 (0.140) [0.140]	−0.003 (0.063) [0.063]	0.000 (0.087) [0.087]	−0.004 (0.119) [0.119]

and Prevention to assess the health and nutritional status of both adults and children in the United States. We study the effect of an average number of alcoholic drinks consumed per day on days when the subject drank alcohol ($\tilde{X}$) on the systolic blood pressure (SBP, mmHg) (Y), adjusting for age (in decade above 50) and body mass index (BMI, kg/m^2) ($\boldsymbol{Z}$). As pointed out in Little and Zhang (2011) and Bartlett *et al.* (2014), it is reasonable to assume that the SBP and BMI are missing completely at random, and thus in our analysis, we only include the

subjects with these two variables fully observed. Among the $n = 2111$ subjects included in the analysis, 720 have missing values for alcohol consumption, and it is reasonable to assume this missingness depends on alcohol consumption itself but is independent of the SBP given alcohol consumption, age, and BMI (Bartlett *et al*., 2014).

The model specifications follow Bartlett *et al*. (2014). Hereafter write $X = \log(\tilde{X} + 1) = \log(\text{no. of drinks} + 1)$. The conditional mean model is

$$E(\text{SBP}|X, \boldsymbol{Z}) = \beta_c + \beta_1 \log(\text{no. of drinks}+1) + \beta_2 \text{BMI} + \beta_3 \text{age} + \beta_4 \text{age}^2,$$

where SBP is centered at 125 mmHg. For the missingness probability $P(R = 1|Y, \boldsymbol{Z})$, a logistic regression is assumed as

$$\text{logit}\{\pi(Y, \boldsymbol{Z}; \boldsymbol{\alpha})\} = \alpha_c + \alpha_1 \text{age} + \alpha_2 \text{BMI} + \alpha_3 \text{SBP} + \alpha_4 \text{SBP}^2.$$

A negative binomial regression is fitted for $f(\tilde{X}|Y, \boldsymbol{Z}, R = 1)$, with all the linear and quadratic terms of age, BMI, and SBP as regressors.

The $\boldsymbol{U}(Y, X, \boldsymbol{Z}; \boldsymbol{\beta})$ is taken to be

$$\begin{pmatrix} 1 \\ \log(\text{no. of drinks} + 1) \\ \text{BMI} \\ \text{age} \\ \text{age}^2 \end{pmatrix} \times (\text{SBP} - \beta_c - \beta_1 \log(\text{no. of drinks}+1) - \beta_2 \text{BMI} - \beta_3 \text{age} - \beta_4 \text{age}^2).$$

We calculate the CC estimator, the ACC estimator, and our proposed estimator using (9), where $\boldsymbol{S}(Y, X, \boldsymbol{Z}; \boldsymbol{\gamma})$ is taken to be the score function for the negative binomial regression model for $f(\tilde{X}|Y, \boldsymbol{Z}, R = 1)$ (eg, Lawless, 1987).

Table 3 contains the results of our data analysis. All methods indicate that alcohol consumption is positively associated with increased SBP adjusting for the other covariates. The same conclusion can be made for BMI. Both the ACC and the proposed methods suggest a significant nonlinear association between age and SBP, while the CC analysis fails to detect the significance. Overall, based on the models considered, the ACC and the proposed methods have similar results and both outperform the CC analysis by providing smaller standard errors.

7 | DISCUSSION

In our proposed method, we jointly estimate the parameter of interest $\boldsymbol{\beta}$ and the nuisance parameter $\boldsymbol{\theta}$ by solving estimating equations altogether using the empirical likelihood method. When the dimension of $\boldsymbol{\theta}$ becomes large, the numerical performance by simultaneously solving all estimating equations may deteriorate. An alternative is to estimate part or all of $\boldsymbol{\theta}$ separately. For example, $\boldsymbol{\alpha}$ in $\pi(Y, \boldsymbol{Z}; \boldsymbol{\alpha})$ and $\boldsymbol{\gamma}$ in $f(\boldsymbol{X}|Y, \boldsymbol{Z}, R = 1; \boldsymbol{\gamma})$ can be separately estimated by maximizing (2) and $\prod_{i=1}^{n} f(\boldsymbol{X}_i|Y_i, \boldsymbol{Z}_i, R_i = 1; \boldsymbol{\gamma})^{R_i}$, respectively, and then $\boldsymbol{\beta}$ can be estimated by the empirical likelihood method using the estimating function

$$\begin{pmatrix} R\boldsymbol{U}(Y, \boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\beta}) \\ \boldsymbol{h}_{use}(Y, \boldsymbol{Z}; \boldsymbol{\beta}, \hat{\theta}) \end{pmatrix},$$

with $\hat{\theta} = (\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\gamma}})$ plugged in. In general, there is no clear efficiency comparison between this alternative method and the CC analysis.

It is noted that our recommended estimator in Section 4 is consistent when working models for both $P(R = 1|Y, \boldsymbol{Z})$ and $f(\boldsymbol{X}|Y, \boldsymbol{Z}, R = 1)$ are correctly specified. Our view of consistency is that it is a very useful theoretical concept, but in practice, models are rarely "correct." Thus we seek methods that are (a) based on

TABLE 3 Analysis results for the NHANES data

	CC		ACC		EL	
	Estimate (SE)	***P* value**	**Estimate (SE)**	***P* value**	**Estimate (SE)**	***P* value**
Intercept	−1.929 (0.798)	0.015	−2.130 (0.741)	0.004	−1.921 (0.745)	0.010
Alcohol[a]	1.267 (0.583)	0.030	1.321 (0.598)	0.027	1.094 (0.550)	0.047
BMI	0.414 (0.080)	<0.001	0.388 (0.066)	<0.001	0.396 (0.062)	<0.001
Age	3.943 (0.261)	<0.001	3.888 (0.198)	<0.001	3.835 (0.227)	<0.001
Age^2	0.265 (0.143)	0.065	0.319 (0.104)	0.002	0.315 (0.107)	0.003

[a]log(number of drinks + 1).

working model assumptions that are checkable from the observed data, (b) are consistent if the working model is correct, and (c) perform well under mild model misspecification. We think that this view is consistent with some other authors'. For example, Seaman and Vansteelandt (2018) have recently given an excellent discussion of "doubly robust" methods for MAR problems, and discuss bias issues and how they are hard to avoid in practice. More numerical investigations are needed to study the performance of the proposed method under model misspecification.

ACKNOWLEDGMENTS

We wish to thank the Editor, the Associate Editor, and two referees for their valuable comments that have helped greatly improve the quality of this work. Funding was provided by the Natural Sciences and Engineering Research Council of Canada under grant RGPIN 8597 to J. F. Lawless.

ORCID

Menglu Che http://orcid.org/0000-0002-0594-5797
Jerald F. Lawless http://orcid.org/0000-0002-3192-0470

REFERENCES

Bartlett, J.W., Carpenter, J.R., Tilling, K. and Vansteelandt, S. (2014) Improving upon the efficiency of complete case analysis when covariates are MNAR. *Biostatistics*, 15, 719–730.

Han, P. (2018) Calibration and multiple robustness when data are missing not at random. *Statistica Sinica*, 28, 1725–1740.

Hastie, T.J. and Tibshirani, R.J. (1990) *Generalized Additive Models*. London: Chapman & Hall.

Horvitz, D.G. and Thompson, D.J. (1952) A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663–685.

Ibrahim, J.G., Lipsitz, S.R. and Chen, M.-H. (1999) 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.

Lawless, J.F. (1987) Negative binomial and mixed Poisson regression. *Canadian Journal of Statistics*, 15, 209–225.

Little, R.J. and Rubin, D.B. (2002) *Statistical Analysis with Missing Data*. New York, NY: John Wiley & Sons.

Little, R.J. and Zhang, N. (2011) Subsample ignorable likelihood for regression analysis with missing data. *Journal of the Royal Statistical Society, Series C*, 60, 591–605.

Miao, W. and Tchetgen Tchetgen, E.J. (2016) On varieties of doubly robust estimators under missingness not at random with a shadow variable. *Biometrika*, 103, 475–482.

Qin, J. and Lawless, J. (1994) Empirical likelihood and general estimating equations. *The Annals of Statistics*, 22, 300–325.

Qin, J., Zhang, B. and Leung, D.H. (2009) Empirical likelihood in missing data problems. *Journal of the American Statistical Association*, 104, 1492–1503.

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.

Rotnitzky, A. and Robins, J. (1997) Analysis of semi-parametric regression models with non-ignorable non-response. *Statistics in Medicine*, 16, 81–102.

Rubin, D. (1987) *Multiple Imputation for Nonresponse in Surveys*. New York, NY: John Wiley & Sons.

Seaman, S.R. and Vansteelandt, S. (2018) Introduction to double robust methods for incomplete data. *Statistical Science*, 33, 184–197.

Tsiatis, A. (2006) *Semiparametric Theory and Missing Data*. New York, NY: Springer.

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

Xie, Y. and Zhang, B. (2017) Empirical likelihood in nonignorable covariate-missing data problems. *The International Journal of Biostatistics*, 13, 1–20.

SUPPORTING INFORMATION

Supporting information referred to in Section 3, together with the R code for our simulation studies, is available at the Biometrics website on Wiley Online Library.

How to cite this article: Che M, Han P, Lawless JF. Improving estimation efficiency for regression with MNAR covariates. *Biometrics*. 2020;76:270–280. https://doi.org/10.1111/biom.13131