



Estimation of logistic regression with covariates missing separately or simultaneously via multiple imputation methods

Shen-Ming Lee¹ · Truong-Nhat Le^{1,2} · Phuoc-Loc Tran³ · Chin-Shang Li⁴

Received: 30 June 2021 / Accepted: 10 June 2022 / Published online: 15 July 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Logistic regression is a standard model in many studies of binary outcome data, and the analysis of missing data in this model is a fascinating topic. Based on the idea of Wang D, Chen SX (2009) Empirical likelihood for estimating equations with missing values. *Ann Stat*, 37:490–517, proposed are two different types of multiple imputation (MI) estimation methods, which each use three empirical conditional distribution functions to generate random values to impute missing data, to estimate the parameters of logistic regression with covariates missing at random (MAR) separately or simultaneously by using the estimating equations of Fay RE (1996) Alternative paradigms for the analysis of imputed survey data. *J Am Stat Assoc*, 91:490–498. The derivation of the two proposed MI estimation methods is under the assumption of MAR separately or simultaneously and exclusively for categorical/discrete data. The two proposed methods are computationally effective, as evidenced by simulation studies. They have a quite similar efficiency and outperform the complete-case, semiparametric inverse probability weighting, validation likelihood, and random forest MI by chained equations methods. Although the two proposed methods are comparable with the joint conditional likelihood (JCL) method, they have

✉ Chin-Shang Li
csli2003@gmail.com

Shen-Ming Lee
smlee@mail.fcu.edu.tw

Truong-Nhat Le
letruongnhat@tdtu.edu.vn

Phuoc-Loc Tran
tploc@ctu.edu.vn

¹ Department of Statistics, Feng Chia University, Taichung, Taiwan, ROC

² Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, Vietnam

³ Department of Mathematics, College of Natural Science, Can Tho University, Can Tho, Vietnam

⁴ School of Nursing, The State University of New York, University at Buffalo, Buffalo, NY, USA

more straightforward calculations and shorter computing times compared to the JCL and MICE methods. Two real data examples are used to illustrate the applicability of the proposed methods.

Keywords Inverse probability weighting · Joint conditional likelihood · Missing at random · Multiple imputation · Validation likelihood

1 Introduction

“Missing data” is a widespread issue in practical data analysis and frequently appear in many areas of science for various reasons, e.g., survey non-response, data collection conditions, expensive or long-term experiments. Rubin (1976) used the missing mechanism concept to formalize the missing value, where the missingness indicators are considered as random variables that are described by a distribution. For the missingness mechanism, there are three main types: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). MCAR means that the appearance of missing values is completely independent of all variables with missing observations and variables with no missing observations. MAR means that the missingness is only related to the variables with no missing observations, but unrelated to the variables with missing observations. MNAR, neither MAR nor MCAR, implies that missing values are related to observed and unobserved values. The impact of missing data on statistical research can lead to biased estimates of parameters, lose information, decrease statistical power, increase standard errors, or weaken generalizability of findings (Dong and Peng 2013). Various methods for dealing with missing values in regression models were proposed; see, e.g., Rubin (1976), Little (1992), Zhao and Lipsitz (1992), Wang et al. (2002), Lee et al. (2012), and Lukusa et al. (2016) for more details.

Logistic regression is often applied in many studies in which researchers would like to investigate the relationship between a binary response variable and covariates (Hosmer et al. 2013). In practice, the analysis of logistic regression with one or more covariates MAR frequently appears, which provides specific challenges. There were some studies on this issue. For instance, Lipsitz et al. (1998) derived a modified conditional logistic regression with covariates MAR. Wang et al. (1997) provided the weighted semiparametric estimation method to investigate the properties of regression parameter estimators when the selection probabilities are estimated by kernel smoothers. Wang et al. (2002) proposed the joint conditional likelihood (JCL) method to estimate the parameters of logistic regression with covariates MAR by combining complete-case (CC) (or validation) data, where the cases without missing observations are included, and non-complete (or non-validation) data that include the cases with missing observations. When both the outcome and covariates in logistic regression are MAR, Lee et al. (2012) presented the two semiparametric estimation methods, validation likelihood (VL) and JCL estimation methods, to estimate the logistic regression parameters. Similarly, Hsieh et al. (2013) applied these approaches to estimate the parameters of logistic regression with the outcome

and covariates MAR separately or simultaneously. Jiang et al. (2020) developed a stochastic approximation version of the EM (SAEM) algorithm, which is based on Metropolis-Hastings sampling, to perform statistical inference for the parameters of logistic regression with missing covariates, and compared their estimators with those of the random forest multiple imputation by chained equations (MICE) method from `mice` package in R (Buuren and Groothuis-Oudshoorn 2011). Tran et al. (2021) recently estimated the parameters of logistic regression with categorical/discrete covariates MAR separately or simultaneously via the JCL estimation method that uses the information from a CC and three non-complete data sets to improve efficiency in estimation.

Based on the results of Tran et al. (2021), although the JCL estimation method outperforms the CC, semiparametric inverse probability weighting (SIPW), and VL estimation methods, its calculations are more complex and, hence, it takes longer computing time. In addition, the estimators of the SAEM approach outperform those of the MICE method, but the computing time of the SAEM approach is longer compared to the MICE method. Moreover, it can be seen from Jiang et al. (2020) that the MICE estimators are underestimators, although the variables are MCAR. Therefore, we are highly motivated to develop other estimation methods for logistic regression with covariates MAR separately or simultaneously that are not only comparable with the JCL and MICE estimation methods in terms of efficiency but also simpler and faster in calculation.

In this work, based on the ideas of Fay (1996), Wang and Chen (2009), and Lee et al. (2016, 2020), we develop two different types of MI methods to estimate the parameters of logistic regression with two covariate vectors MAR separately or simultaneously under the assumption that all covariates and surrogates are categorical/discrete. We also compare the two proposed MI methods with the JCL method of Tran et al. (2021) and MICE method in terms of efficiency in estimation and computing time. Our proposed MI methods are two-step procedures based on the suggestion of Fay (1996) that makes the calculation process simpler and faster than the three-step procedure in Rubin (1987). Firstly, each type MI estimation method uses three empirical conditional distribution functions (CDFs) (Wang and Chen 2009) to generate random values. The first type MI (MI1) method uses only the CC data. The second type MI (MI2) method uses both the CC and non-complete data to generate random values to impute missing data. Secondly, solve the estimating equations to obtain estimates of the logistic regression parameters (Fay 1996). The estimating equations are more convenient in practice because they are solved only once rather than combining M estimating equations, where M is the number of imputations, to get the pooled estimates as done in Rubin (1987), such that it shortens computing time. The formulas of Lee et al. (2016, 2020) are applied to estimate the variances of the two proposed MI estimators to improve efficiency in estimation.

Section 2 presents the assumptions and notations used throughout this work. Section 3 reviews the SIPW, VL, JCL, and MICE estimation methods. Two different types of MI estimation methods are proposed in Sect. 4. In Sect. 5, the finite-sample performances of the proposed methods are investigated by conducting extensive simulations under various settings. Two real data sets are used to demonstrate the practical use of the proposed methods. Section 6 presents some discussions and conclusions.

2 Assumptions and notations

Let Y be a binary outcome variable denoting whether an event of interest occurs, where $Y = 1$ if event occurs and $Y = 0$ otherwise. Suppose that $X_1 = (X_{11}, X_{12}, \dots, X_{1r_1})^T$ is a vector of r_1 categorical/discrete covariates, and $X_2 = (X_{2r_1+1}, X_{2r_1+2}, \dots, X_{2p})^T$ is a vector of r_2 categorical/discrete covariates, where $p = r_1 + r_2$. It is assumed that X_1 and X_2 may be missing separately or simultaneously. In this work, we do not consider the case where some covariates in X_1 and X_2 have missing observations. Thus, when it is said that X_s , $s = 1, 2$, is missing, it means all covariates in X_s have missing observations simultaneously. Assume that $\mathbf{Z} = (Z_1, Z_2, \dots, Z_q)^T$ a vector of q categorical/discrete covariates that are always observed. Let $\mathbf{X} = (X_1^T, X_2^T)^T$, $\mathcal{X} = (1, \mathbf{X}^T, \mathbf{Z}^T)^T$, and $\{(Y_i, \mathcal{X}_i) : i = 1, 2, \dots, n\}$ be a random sample. Assume that \mathbf{X} is MAR. The logistic regression model is considered as follows:

$$P(Y_i = 1 | X_i, Z_i) = H(\beta_0 + \beta_1^T X_{1i} + \beta_2^T X_{2i} + \beta_3^T Z_i) = H(\beta^T \mathcal{X}_i), \quad i = 1, 2, \dots, n, \tag{1}$$

where $H(u) = \{1 + \exp(-u)\}^{-1}$ and $\beta = (\beta_0, \beta_1^T, \beta_2^T, \beta_3^T)^T$ is a vector of parameters associated with \mathcal{X}_i . The main goal is to estimate β when some of the X_i s are MAR. Let δ_{ij} , $j = 1, 2, 3, 4$, denote the missingness statuses of $X_i = (X_{1i}^T, X_{2i}^T)^T$ in which $\delta_{i1} = 1$ if both X_{1i} and X_{2i} are observed; 0 otherwise; $\delta_{i2} = 1$ if X_{1i} is missing and X_{2i} is observed; 0 otherwise; $\delta_{i3} = 1$ if X_{1i} is observed and X_{2i} is missing; 0 otherwise; $\delta_{i4} = 1$ if both X_{1i} and X_{2i} are missing; 0 otherwise. Some studies used a surrogate variable that can be measured or is measured easily that is used instead of a variable that cannot be measured or is measured difficultly, so it can be used for an MAR variable in a model to improve the information and, hence, enhance the performance of estimation and prediction. See, e.g., Wang et al. (1997, 2002), Hsieh et al. (2010, 2013), and Lee et al. (2011, 2012, 2020) for more details. We also consider the possibility of categorical/discrete surrogate vectors W_1 and W_2 for X_1 and X_2 , respectively, such that W_1 and W_2 are dependent on X_1 and X_2 , respectively, and independent of Y given \mathbf{X} and \mathbf{Z} . That is, there are correlations between W_1 and X_1 and W_2 and X_2 , respectively. Hence, we can have the logistic regression model $P(Y_i = 1 | X_{1i}, X_{2i}, Z_i, W_{1i}, W_{2i}) = P(Y_i = 1 | X_{1i}, X_{2i}, Z_i) = H(\beta_0 + \beta_1^T X_{1i} + \beta_2^T X_{2i} + \beta_3^T Z_i)$ as given in (1). Let $\mathbf{W} = (W_1^T, W_2^T)^T$ and $\mathbf{V}_i = (Z_i^T, \mathbf{W}_i^T)^T$, $i = 1, \dots, n$. The CC data set ($\delta_{i1} = 1$) consists of (Y_i, X_i, V_i) , and the three non-complete data sets include (Y_i, X_{2i}, V_i) , (Y_i, X_{1i}, V_i) , and (Y_i, V_i) , respectively, when δ_{i2} , δ_{i3} , and δ_{i4} are equal to 1. Under the assumption of MAR mechanism (Rubin 1976) of X_1 and X_2 , the selection probability model is

$$P(\delta_{ij} = 1 | Y_i, X_{1i}, X_{2i}, Z_i, \mathbf{W}_i) = \pi_j(Y_i, Z_i, \mathbf{W}_i) = \pi_j(Y_i, V_i), \quad j = 1, 2, 3, 4, \tag{2}$$

with $\sum_{j=1}^4 \pi_j(Y_i, V_i) = 1$. $\pi_j(Y_i, V_i)$ s are the unknown nuisance parameters and need to be estimated. In this study, under the assumption that V_i s are categorical/discrete vectors, the nonparametric estimators of $\pi_j(Y_i, V_i)$ are given as follows:

$$\hat{\pi}_j(Y_i, V_i) = \frac{\sum_{k=1}^n \delta_{kj} I(Y_k = Y_i, V_k = V_i)}{\sum_{s=1}^n I(Y_s = Y_i, V_s = V_i)}, \quad j = 1, 2, 3, 4, \tag{3}$$

where $I(\cdot)$ is an indicator function. It is noticed that in this work the requirement of categorical/discrete covariates is for mathematical derivation purposes. $\hat{\pi}_1(Y_i, \mathbf{V}_i)$ s are used as weights for the SIPW and VL estimation methods and $\hat{\pi}_j(Y_i, \mathbf{V}_i)$ s, $j = 1, 2, 3, 4$, are used as weights for modification of conditional probabilities for the JCL estimation method in Sect. 3. If covariates are continuous, one can use the kernel estimation approach by using the arguments of Wang and Wang (2001).

3 Review of estimation methods

This section briefly reviews the four famous estimation methods, SIPW, VL, JCL, and MICE methods, for logistic regression with covariates MAR separately or simultaneously. These methods are also used to compare their estimation performance with the proposed approaches in the following sections.

3.1 SIPW estimation method

Horvitz and Thompson (1952) proposed the weighted estimator that uses inverse probability weighting (IPW) to reduce biased estimation and has become known as the H-T estimator. When the selection probabilities are known, Zhao and Lipsitz (1992) extended the H-T estimator to propose the IPW estimator to improve efficiency in estimation. This approach is, however, limited in practice because the selection probabilities are usually unknown. Therefore, some authors, e.g., Wang et al. (1997) and Wang and Wang (2001), suggested the SIPW approach that uses the nonparametric estimators of the unknown selection probabilities as weighted inverse terms. See, e.g., Hsieh et al. (2010) and Lee et al. (2012) for further details. Considering the logistic regression model (1) when X_1 and X_2 are MAR separately or simultaneously, one can obtain the SIPW estimator $\hat{\beta}_W$ of β by solving the following estimating equations:

$$U_W(\beta, \hat{\pi}_1) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_{i1}}{\hat{\pi}_1(Y_i, \mathbf{V}_i)} \mathcal{X}_i(Y_i - H(\beta^T \mathcal{X}_i)) = \mathbf{0}, \tag{4}$$

where $\hat{\pi}_1 = (\hat{\pi}_{11}, \dots, \hat{\pi}_{1n})$ for $\hat{\pi}_{1i} = \hat{\pi}_1(Y_i, \mathbf{V}_i)$, given in (3), being the estimator of $\pi_{1i} = \pi_1(Y_i, \mathbf{V}_i)$ defined in (2).

3.2 VL estimation method

Breslow and Cain (1988) proposed the conditional maximum likelihood (ML) approach to estimate the parameters of logistic regression for two-stage case-control data. They showed that their estimator of β is not only consistent and asymptotically a normal distribution but also is useful when information of covariates is missing for a large part of the sample. Wang et al. (2002), Lee et al. (2012), and Hsieh et al. (2013) applied this approach to provide the VL estimation method, which uses the CC data, to solve the issue of missing data in logistic regression. When X_1 and X_2 are MAR

separately or simultaneously, Tran et al. (2021) estimated β by using the following estimating equations:

$$\widehat{U}_V(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_{i1} \mathcal{X}_i(Y_i - \widehat{H}_1(X_{1i}, X_{2i}, V_i; \beta)) = \mathbf{0}, \tag{5}$$

where

$$\widehat{H}_1(X_{1i}, X_{2i}, V_i; \beta) = H\left(\beta_0 + \beta_1^T X_{1i} + \beta_2^T X_{2i} + \beta_3^T Z_i + \ln \frac{\widehat{\pi}_1(1, V_i)}{\widehat{\pi}_1(0, V_i)}\right) \tag{6}$$

and $\widehat{\pi}_1(Y_i, V_i)$ is given in (3). One can solve $\widehat{U}_V(\beta) = \mathbf{0}$ to obtain the VL estimator $\widehat{\beta}_V$ of β .

3.3 JCL estimation method

Both the SIPW and VL approaches that use only the CC data set ($\delta_1 = 1$) may not maximize efficiency in estimation. Therefore, to overcome this drawback, Wang et al. (2002) proposed the JCL estimation method that combines the CC and non-complete data; see, e.g., Lee et al. (2012) and Hsieh et al. (2013) for more details. Tran et al. (2021) proposed the JCL method to estimate the parameters of logistic regression with covariates MAR separately or simultaneously.

When X_1 and X_2 are MAR separately or simultaneously, one can obtain the JCL estimator $\widehat{\beta}_J = (\widehat{\beta}_0, \widehat{\beta}_{J1}^T, \widehat{\beta}_{J2}^T, \widehat{\beta}_{J3}^T)^T$ of β in the logistic regression model (1) by solving the following estimating equations:

$$\begin{aligned} \widehat{U}_J(\beta) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \{ \delta_{i1} \mathcal{X}_i(Y_i - \widehat{H}_1(X_{1i}, X_{2i}, V_i; \beta)) + \delta_{i2} \widehat{\mathcal{A}}(X_{2i}, V_i; \beta_1)(Y_i - \widehat{H}_2(X_{2i}, V_i; \beta)) \\ &\quad + \delta_{i3} \widehat{\mathcal{B}}(X_{1i}, V_i; \beta_2)(Y_i - \widehat{H}_3(X_{1i}, V_i; \beta)) + \delta_{i4} \widehat{\mathcal{T}}(V_i; \beta_1, \beta_2)(Y_i - \widehat{H}_4(V_i; \beta)) \} \\ &= \mathbf{0}, \end{aligned} \tag{7}$$

where $\widehat{H}_1(X_{1i}, X_{2i}, V_i; \beta)$ is defined in (6), and

$$\widehat{H}_2(X_{2i}, V_i; \beta) = H\left(\beta_0 + \beta_2^T X_{2i} + \beta_3^T Z_i + \widehat{R}_2(X_{2i}, V_i; \beta_1) + \ln \frac{\widehat{\pi}_2(1, V_i)}{\widehat{\pi}_2(0, V_i)}\right), \tag{8}$$

$$\widehat{H}_3(X_{1i}, V_i; \beta) = H\left(\beta_0 + \beta_1^T X_{1i} + \beta_3^T Z_i + \widehat{R}_3(X_{1i}, V_i; \beta_2) + \ln \frac{\widehat{\pi}_3(1, V_i)}{\widehat{\pi}_3(0, V_i)}\right), \tag{9}$$

$$\widehat{H}_4(V_i; \beta) = H\left(\beta_0 + \beta_3^T Z_i + \widehat{R}_4(V_i; \beta_1, \beta_2) + \ln \frac{\widehat{\pi}_4(1, V_i)}{\widehat{\pi}_4(0, V_i)}\right), \tag{10}$$

$$\widehat{A}(X_{2i}, V_i; \beta_1) = \left(1, \widehat{R}_2^{(\beta_1)}(X_{2i}, V_i; \beta_1), X_{2i}^T, Z_i^T \right)^T, \tag{11}$$

$$\widehat{B}(X_{1i}, V_i; \beta_2) = \left(1, X_{1i}^T, \widehat{R}_3^{(\beta_2)}(X_{1i}, V_i; \beta_2), Z_i^T \right)^T, \tag{12}$$

$$\widehat{T}(V_i; \beta_1, \beta_2) = \left(1, \widehat{R}_4^{(\beta_1)}(V_i; \beta_1, \beta_2), \widehat{R}_4^{(\beta_2)}(V_i; \beta_1, \beta_2), Z_i^T \right)^T, \tag{13}$$

for

$$\widehat{R}_2(X_{2i}, V_i; \beta_1) = \ln \frac{\sum_{j=1}^n \delta_{j1} e^{\beta_1^T X_{2j}} I(Y_j = 0, X_{2j} = X_{2i}, V_j = V_i)}{\sum_{k=1}^n \delta_{k1} I(Y_k = 0, X_{2k} = X_{2i}, V_k = V_i)}, \tag{14}$$

$$\widehat{R}_3(X_{1i}, V_i; \beta_2) = \ln \frac{\sum_{j=1}^n \delta_{j1} e^{\beta_2^T X_{2j}} I(Y_j = 0, X_{1j} = X_{1i}, V_j = V_i)}{\sum_{k=1}^n \delta_{k1} I(Y_k = 0, X_{1k} = X_{1i}, V_k = V_i)}, \tag{15}$$

$$\widehat{R}_4(V_i; \beta_1, \beta_2) = \ln \frac{\sum_{j=1}^n \delta_{j1} e^{\beta_1^T X_{1j} + \beta_2^T X_{2j}} I(Y_j = 0, V_j = V_i)}{\sum_{k=1}^n \delta_{k1} I(Y_k = 0, V_k = V_i)}, \tag{16}$$

and

$$\begin{aligned} \widehat{R}_2^{(\beta_1)}(X_{2i}, V_i; \beta_1) &= \frac{\partial}{\partial \beta_1^T} \widehat{R}_2(X_{2i}, V_i; \beta_1), & \widehat{R}_3^{(\beta_2)}(X_{1i}, V_i; \beta_2) &= \frac{\partial}{\partial \beta_2^T} \widehat{R}_3(X_{1i}, V_i; \beta_2), \\ \widehat{R}_4^{(\beta_1)}(V_i; \beta_1, \beta_2) &= \frac{\partial}{\partial \beta_1^T} \widehat{R}_4(V_i; \beta_1, \beta_2), & \widehat{R}_4^{(\beta_2)}(V_i; \beta_1, \beta_2) &= \frac{\partial}{\partial \beta_2^T} \widehat{R}_4(V_i; \beta_1, \beta_2). \end{aligned}$$

3.4 MICE estimation method

MI (Rubin 1987, 1996) is a simple and powerful method for dealing with missing data and is available in several commonly used statistical software and packages. One can use the MI approaches to generate fully imputed (“completed”) data sets by retaining the observed values and replacing the missing data with plausible values from an imputation model. There are various extensions of the MI method of Rubin (1987) in which their differences are mostly in the structural assumption of the imputation model. See, e.g., Rubin (1987, 1996), Rubin and Schenker (1986), Fay (1996), and Pahel et al. (2011) for details. Their procedures often follow the three basic steps, shown in Figure 1A: Step 1) *Imputation*: impute missing values M times to obtain M imputed (“completed”) data sets; Step 2) *Analysis*: analyze each of the M imputed (“completed”) data sets by using the chosen statistical methods; Step 3) *Pool*: combine the M analysis results from Step 2 into one result by using the formula of Rubin (1987). The MI methods are

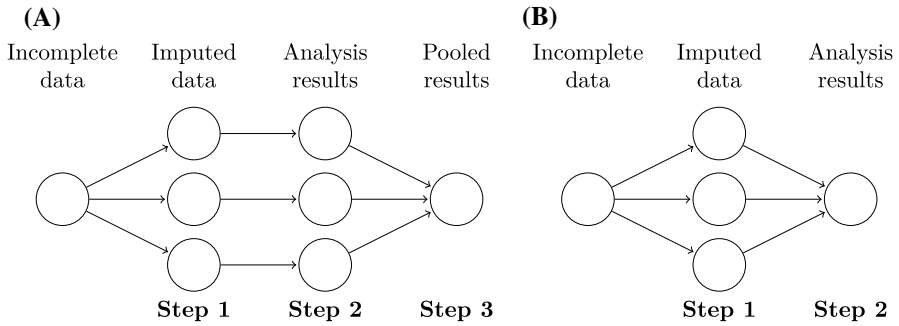


Fig. 1 The main steps of multiple imputation. **A** Rubin's type. **B** Fay's type

almost different in the way of their imputed data to fill missing values (Step 1), e.g., mean, regression, hot deck, cold deck, principal component, and chained-equation imputation methods (Little and Rubin 2019).

The `mice` package (Buuren and Groothuis-Oudshoorn 2011) is one of the useful and famous packages that uses MI by chained equations, also known as fully conditional specification and sequential regression MI, in Step 1 to impute values for missing data. Specifically, assume $\mathbf{u} = (u_1, u_2, \dots, u_k)$ is a vector of k variables with missing observations. Firstly, all missing values of u_i are filled in by using basic random sampling and replacement with observed values themselves. Secondly, construct a regression model of u_1 on u_2, \dots, u_k . Simulated draws from the corresponding posterior predictive distribution of u_1 are used to replace missing values of u_1 . Then, generate a regression model of u_2 based on the u_1 variable with imputed values and u_3, \dots, u_k . Again, simulated draws from the corresponding posterior predictive distribution of u_2 are used to replace missing values of u_2 . Repeat the process for each of u_3, \dots, u_k . This procedure is called a “cycle”. Repeat this cycle several times (e.g., 10 or 20) to create a first imputed (“complete”) data set. Finally, repeat the procedure M times to obtain M imputed (“completed”) data sets. Steps 2 and 3 of this MI method follow the rule of Rubin (1987).

In summary, this section has introduced four estimation methods used for the benchmark comparisons in this work and the convenience of reference. Although the simulation results of Tran et al. (2021) showed that the JCL method outperforms the CC, VL, and SIPW methods, it is pretty hard to obtain the JCL estimates of the logistic regression parameters because of its complex calculations and long computing time. In addition, the MICE approach can be applied to the case of both continuous and categorical/discrete data and any missingness mechanism (White et al. 2011), but its estimators for logistic regression with missing covariates underestimate the parameters, although their standard errors were small (Jiang et al. 2020). Moreover, the procedure of this approach may take much time to compute. Therefore, in Sect. 4, we propose two different types of MI methods that not only provide efficient estimation comparable with the JCL and MICE approaches but also have more straightforward computations to shorten computing time.

4 Proposed MI estimation methods

This section introduces two different types of MI methods based on the ideas of Fay (1996), Wang and Chen (2009), and Lee et al. (2016, 2020). The procedure of these two proposed MI methods has two steps (Fay’s type, Figure 1B): Step 1) Impute values for non-complete data by using the empirical CDFs of missing values given the observed data as done in Wang and Chen (2009) to obtain M imputed (“completed”) data sets; Step 2) Solve the estimating equations only one time to obtain the estimates of the logistic regression parameters. The estimated variances of the MI estimators are then obtained by using the formulas of Lee et al. (2016, 2020). The main difference of these two MI methods is in the way of missing data imputation in Step 1. As mentioned in Sect. 1, this procedure is more convenient in practice and saves computing time. The details of these two different types of MI methods are stated in the following sections.

4.1 Type 1 MI (MI1) method

Let $F_{X_{1i}}(x_1|X_{2i}, Y_i, \mathbf{V}_i)$, $F_{X_{2i}}(x_2|X_{1i}, Y_i, \mathbf{V}_i)$, and $F_{X_i}(\mathbf{x}|Y_i, \mathbf{V}_i)$, where $\mathbf{x} = (x_1^T, x_2^T)^T$, be the CDFs of X_{1i} given $(X_{2i}, Y_i, \mathbf{V}_i)$, X_{2i} given $(X_{1i}, Y_i, \mathbf{V}_i)$, and $X_i = (X_{1i}^T, X_{2i}^T)^T$ given (Y_i, \mathbf{V}_i) , respectively. To build the MI1 method, we consider the following empirical CDFs of X_{1i} given $(X_{2i}, Y_i, \mathbf{V}_i)$, X_{2i} given $(X_{1i}, Y_i, \mathbf{V}_i)$, and X_i given (Y_i, \mathbf{V}_i) :

$$\begin{aligned} \tilde{F}_{X_{1i}}(x_1|X_{2i}, Y_i, \mathbf{V}_i) &= \sum_{k=1}^n \left(\frac{\delta_{k1} I(Y_k = Y_i, X_{2k} = X_{2i}, \mathbf{V}_k = \mathbf{V}_i)}{\sum_{s=1}^n \delta_{s1} I(Y_s = Y_i, X_{2s} = X_{2i}, \mathbf{V}_s = \mathbf{V}_i)} \right) I(X_{1k} \leq x_1), \\ \tilde{F}_{X_{2i}}(x_2|X_{1i}, Y_i, \mathbf{V}_i) &= \sum_{k=1}^n \left(\frac{\delta_{k1} I(Y_k = Y_i, X_{1k} = X_{1i}, \mathbf{V}_k = \mathbf{V}_i)}{\sum_{s=1}^n \delta_{s1} I(Y_s = Y_i, X_{1s} = X_{1i}, \mathbf{V}_s = \mathbf{V}_i)} \right) I(X_{2k} \leq x_2), \\ \tilde{F}_{X_i}(\mathbf{x}|Y_i, \mathbf{V}_i) &= \sum_{k=1}^n \left(\frac{\delta_{k1} I(Y_k = Y_i, \mathbf{V}_k = \mathbf{V}_i)}{\sum_{s=1}^n \delta_{s1} I(Y_s = Y_i, \mathbf{V}_s = \mathbf{V}_i)} \right) I(\mathbf{X}_k \leq \mathbf{x}), \end{aligned} \tag{17}$$

respectively. When X_{1i} and X_{2i} are MAR separately or simultaneously, their missing values are imputed several times by random values generated from $\tilde{F}_{X_{1i}}(x_1|X_{2i}, Y_i, \mathbf{V}_i)$, $\tilde{F}_{X_{2i}}(x_2|X_{1i}, Y_i, \mathbf{V}_i)$, and $\tilde{F}_{X_i}(\mathbf{x}|Y_i, \mathbf{V}_i)$. The MI1 procedure is summarized as follows:

Step 1. Imputation: Generate the v th imputed (“completed”) data ($v = 1, 2, \dots, M$) based on the missingness status of $X_i = (X_{1i}^T, X_{2i}^T)^T$, $i = 1, 2, \dots, n$.

- i) If $\delta_{i1} = 1$ (X_{1i} and X_{2i} are observed), keep the values of X_{1i} and X_{2i} . Set $\mathcal{X}_i = (1, X_{1i}^T, X_{2i}^T, \mathbf{Z}_i^T)^T$ for all v .

- ii) If $\delta_{i2} = 1$ (X_{1i} is missing and X_{2i} is observed), keep the values of X_{2i} and generate \tilde{X}_{1iv} from $\tilde{F}_{X_{1i}}(x_1|X_{2i}, Y_i, V_i)$ to fill the missing values of X_{1i} . Define $\tilde{\mathcal{X}}_{2iv} = (1, \tilde{X}_{1iv}^T, X_{2i}^T, Z_i^T)^T$.
- iii) If $\delta_{i3} = 1$ (X_{1i} is observed and X_{2i} is missing), keep the values of X_{1i} and generate \tilde{X}_{2iv} from $\tilde{F}_{X_{2i}}(x_2|X_{1i}, Y_i, V_i)$ to fill the missing values of X_{2i} . Define $\tilde{\mathcal{X}}_{3iv} = (1, X_{1i}^T, \tilde{X}_{2iv}^T, Z_i^T)^T$.
- iv) If $\delta_{i4} = 1$ (both X_{1i} and X_{2i} are missing), generate \tilde{X}_{1iv} and \tilde{X}_{2iv} from $\tilde{F}_{X_i}(x|Y_i, V_i)$ to fill the missing values of X_{1i} and X_{2i} . Define $\tilde{\mathcal{X}}_{4iv} = (1, \tilde{X}_{1iv}^T, \tilde{X}_{2iv}^T, Z_i^T)^T$.

Step 2. Analysis: Solve the following estimating equations:

$$U_{M1}(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\delta_{i1} S_i(\beta) + \delta_{i2} \tilde{S}_{2i}(\beta) + \delta_{i3} \tilde{S}_{3i}(\beta) + \delta_{i4} \tilde{S}_{4i}(\beta) \right) = \mathbf{0}, \quad (18)$$

where $S_i(\beta) = \mathcal{X}_i(Y_i - H(\beta^T \mathcal{X}_i))$, $\tilde{S}_{ki}(\beta) = M^{-1} \sum_{v=1}^M \tilde{\mathcal{X}}_{kiv}(Y_i - H(\beta^T \tilde{\mathcal{X}}_{kiv}))$, $k = 2, 3, 4$, to obtain the MI1 estimator, $\hat{\beta}_{M1}$, of β . Next, calculate the estimated variance of $\hat{\beta}_{M1}$, $\widehat{\text{Var}}(\hat{\beta}_{M1})$, by the formula of Lee et al. (2016), which is also a Rubin-type estimated variance (Rubin 1987), as follows:

$$G_{M1}^{-1}(\hat{\beta}_{M1}) \left\{ \frac{1}{M} \sum_{v=1}^m \sum_{i=1}^n (\tilde{U}_{vi}(\hat{\beta}_{M1}))^{\otimes 2} + \left(1 + \frac{1}{M}\right) \frac{\sum_{v=1}^m (\tilde{U}_v(\hat{\beta}_{M1}))^{\otimes 2}}{M-1} \right\} (G_{M1}^{-1}(\hat{\beta}_{M1}))^T, \quad (19)$$

where $G_{M1}^{-1}(\beta)$ is the gradient of $-M^{-1} \sum_{v=1}^M \tilde{U}_v(\beta) = -U_{M1}(\beta)$, for

$$\begin{aligned} \tilde{U}_{vi}(\beta) &= \frac{1}{\sqrt{n}} \left(\delta_{i1} S_i(\beta) + \delta_{i2} \tilde{S}_{2i}(\beta) + \delta_{i3} \tilde{S}_{3i}(\beta) + \delta_{i4} \tilde{S}_{4i}(\beta) \right), \\ \tilde{U}_v(\beta) &= \sum_{i=1}^n \tilde{U}_{vi}(\beta). \end{aligned}$$

Notice that in Step 1, the imputed data sets are generated by random sampling from the empirical CDFs based on the CC data. The indicator variables δ_{ij} , $j = 1, 2, 3, 4$, $i = 1, 2, \dots, n$, are used to identify exactly the partitioned covariate vector without missing observations that are used as the information for the empirical CDFs. For example, when $\delta_{i2} = 1$ ($\delta_{i3} = 1$), the condition from the observed X_{2i} (X_{1i}), Y_i , and V_i is used to create a set of values for the missing values of X_{1i} (X_{2i}); when $\delta_{i4} = 1$, both X_{1i} and X_{2i} are missing, the condition from only Y_i and V_i is used to create a set of values for the missing values of X_{1i} and X_{2i} . In other words, the procedure uses the exact and more information from the covariate vector without missing observations for missing data imputation, and,

hence, estimation may be more efficient. In addition, because this method only has two steps and solves the estimating equations one time, it can shorten computing time.

4.2 Type 2 MI (MI2) method

The MI2 method is quite similar to the MI1 method except for the formulas of empirical CDFs, which are used to generate values to fill missing data. Consider the empirical CDFs of X_{1i} given (X_{2i}, Y_i, V_i) , X_{2i} given (X_{1i}, Y_i, V_i) , and X_i given (Y_i, V_i) as follows:

$$\begin{aligned} \tilde{F}_{X_{1i}}(x_1|Y_i, V_i) &= \sum_{k=1}^n \left(\frac{(\delta_{k1} + \delta_{k3})I(Y_k = Y_i, V_k = V_i)}{\sum_{s=1}^n (\delta_{s1} + \delta_{s3})I(Y_s = Y_i, V_s = V_i)} \right) I(X_{1k} \leq x_1), \\ \tilde{F}_{X_{2i}}(x_2|Y_i, V_i) &= \sum_{k=1}^n \left(\frac{(\delta_{k1} + \delta_{k2})I(Y_k = Y_i, V_k = V_i)}{\sum_{s=1}^n (\delta_{s1} + \delta_{s2})I(Y_s = Y_i, V_s = V_i)} \right) I(X_{2k} \leq x_2), \\ \tilde{F}_{X_i}(\mathbf{x}|Y_i, V_i) &= \sum_{k=1}^n \left(\frac{\delta_{k1}I(Y_k = Y_i, V_k = V_i)}{\sum_{s=1}^n \delta_{s1}I(Y_s = Y_i, V_s = V_i)} \right) I(\mathbf{X}_k \leq \mathbf{x}), \end{aligned}$$

respectively. When X_{1i} and X_{2i} are MAR separately or simultaneously, their missing values are imputed several times by random observations generated from $\tilde{F}_{X_{1i}}(x_1|Y_i, V_i)$, $\tilde{F}_{X_{2i}}(x_2|Y_i, V_i)$, and $\tilde{F}_{X_i}(\mathbf{x}|Y_i, V_i)$ according to the missingness statuses of X_{1i} and X_{2i} . The procedure of the MI2 method is stated as follows:

Step 1. Imputation: Generate the v th imputed (“completed”) data ($v = 1, 2, \dots, M$) according to the missingness status of $X_i = (X_{1i}^T, X_{2i}^T)^T$, $i = 1, 2, \dots, n$.

- i) If $\delta_{i1} = 1$, keep the values of X_{1i} and X_{2i} . Set $\mathcal{X}_i = (1, X_{1i}^T, X_{2i}^T, \mathbf{Z}_i^T)^T$ for all v .
- ii) If $\delta_{i2} = 1$, keep the values of X_{2i} and generate \tilde{X}_{1iv} from $\tilde{F}_{X_{1i}}(x_1|X_{2i}, Y_i, V_i)$ to fill the missing values of X_{1i} . Define $\tilde{\mathcal{X}}_{2iv} = (1, \tilde{X}_{1iv}^T, X_{2i}^T, \mathbf{Z}_i^T)^T$.
- iii) If $\delta_{i3} = 1$, keep the values of X_{1i} and generate \tilde{X}_{2iv} from $\tilde{F}_{X_{2i}}(x_2|X_{1i}, Y_i, V_i)$ to fill the missing values of X_{2i} . Define $\tilde{\mathcal{X}}_{3iv} = (1, X_{1i}^T, \tilde{X}_{2iv}^T, \mathbf{Z}_i^T)^T$.
- iv) If $\delta_{i4} = 1$, generate \tilde{X}_{1iv} and \tilde{X}_{2iv} from $\tilde{F}_{X_i}(\mathbf{x}|Y_i, V_i)$ to fill the missing values of X_{1i} and X_{2i} . Define $\tilde{\mathcal{X}}_{4iv} = (1, \tilde{X}_{1iv}^T, \tilde{X}_{2iv}^T, \mathbf{Z}_i^T)^T$.

Step 2. Analysis: Solve the following estimating equations:

$$U_{M2}(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\delta_{i1} S_i(\beta) + \delta_{i2} \tilde{S}_{2i}(\beta) + \delta_{i3} \tilde{S}_{3i}(\beta) + \delta_{i4} \tilde{S}_{4i}(\beta) \right) = \mathbf{0}, \quad (20)$$

where $S_i(\beta) = \mathcal{X}_i(Y_i - H(\beta^T \mathcal{X}_i))$, $\tilde{S}_{ki}(\beta) = M^{-1} \sum_{v=1}^M \tilde{\mathcal{X}}_{kiv}(Y_i - H(\beta^T \tilde{\mathcal{X}}_{kiv}))$, $k = 2, 3, 4$, to obtain the MI2 estimator, $\hat{\beta}_{M2}$, of β . Then, calculate the estimated variance of $\hat{\beta}_{M2}$, $\widehat{\text{Var}}(\hat{\beta}_{M2})$, by the following formulas:

$$G_{M2}^{-1}(\hat{\beta}_{M2}) \left\{ \frac{1}{M} \sum_{v=1}^m \sum_{i=1}^n (\tilde{U}_{vi}(\hat{\beta}_{M2}))^{\otimes 2} + \left(1 + \frac{1}{M} \right) \frac{\sum_{v=1}^m (\tilde{U}_v(\hat{\beta}_{M2}))^{\otimes 2}}{M-1} \right\} (G_{M2}^{-1}(\hat{\beta}_{M2}))^T, \quad (21)$$

where $G_{M2}^{-1}(\beta)$ is the gradient of $-M^{-1} \sum_{v=1}^M \tilde{U}_v(\beta) = -U_{M2}(\beta)$, for

$$\begin{aligned} \tilde{U}_{vi}(\beta) &= \frac{1}{\sqrt{n}} \left(\delta_{i1} S_i(\beta) + \delta_{i2} \tilde{S}_{2i}(\beta) + \delta_{i3} \tilde{S}_{3i}(\beta) + \delta_{i4} \tilde{S}_{4i}(\beta) \right), \\ \tilde{U}_v(\beta) &= \sum_{i=1}^n \tilde{U}_{vi}(\beta). \end{aligned}$$

Notice that each of the two proposed MI methods uses three empirical CDFs of Wang and Chen (2009) to generate random values for X_1 and X_2 MAR separately or simultaneously. However, there is a slight difference between these two MI approaches in how information is obtained from the observed data to replace the missing data in the first and second empirical CDFs, while the third empirical CDFs are the same. The MI1 method uses only $\delta_{i1} = 1$, but the information of observed X_{2i} is given in the indicator function in $\tilde{F}_{X_{1i}}(x_1 | X_{2i}, Y_i, V_i)$, and the information of observed X_{1i} is given in the indicator function in $\tilde{F}_{X_{2i}}(x_2 | X_{1i}, Y_i, V_i)$. The MI2 method, however, combines $\delta_{i1} = 1$ and $\delta_{i3} = 1$ in $\tilde{F}_{X_{1i}}(x_1 | Y_i, V_i)$, and $\delta_{i1} = 1$ and $\delta_{i2} = 1$ in $\tilde{F}_{X_{2i}}(x_2 | Y_i, V_i)$, but the information of observed X_{1i} or X_{2i} is not given in the indicator functions $I(Y_k = Y_i, V_k = V_i)$ s. Moreover, by the conditions inside the indicator functions in the empirical CDFs, the MI2 method can be applied to the case of both the continuous and categorical/discrete X data, while the MI1 method can only be applied to the case of the categorical/discrete X data, which is its limitation in practice. Despite their differences in methodology, both the techniques aim to improve information for missing data imputation and maximize estimation efficiencies. The next section investigates the finite-sample performances of the proposed methods compared to the CC, SIPW, VL, JCL, and MICE estimation methods via extensive simulations.

5 Simulation and real data studies

5.1 Simulation studies

Monte Carlo simulations were conducted to examine the finite-sample performances of the following estimators:

- (1) $\hat{\beta}_F$: full data ML estimator used as a benchmark for comparisons
- (2) $\hat{\beta}_C$: CC estimator
- (3) $\hat{\beta}_W$: SIPW estimator that is the solution of $U_W(\beta, \hat{\pi}_1) = \mathbf{0}$ in (4)
- (4) $\hat{\beta}_V$: VL estimator that is the solution of $\hat{U}_V(\beta) = \mathbf{0}$ in (5)
- (5) $\hat{\beta}_J$: JCL estimator that is the solution of $\hat{U}_J(\beta) = \mathbf{0}$ in (7)
- (6) $\hat{\beta}_{M1}$: MI1 estimator that is the solution of $U_{M1}(\beta) = \mathbf{0}$ in (18)
- (7) $\hat{\beta}_{M2}$: MI2 estimator that is the solution of $U_{M2}(\beta) = \mathbf{0}$ in (20)
- (8) $\hat{\beta}_{ME}$: MICE estimator from `mice` package in R, used as a benchmark for comparisons.

To evaluate the estimation performances of the proposed methods as well as compare with the other estimation methods under various situations, we were motivated to construct five scenarios, where the two univariates X_1 and X_2 were uncorrelated in Scenarios 1-4, and correlated in Scenario 5. Specifically, Scenario 1 studied the impact of the sample size on the performance of all the estimation methods under the same selection probabilities. Three sets of selection probabilities were considered in Scenario 2 to examine the influence of the missing rates on the performances of the estimation methods. The aim of Scenario 3 was the same as Scenario 2 except only changing the value of β to know whether the efficiencies of the estimators were altered or not under different logistic regression models. In Scenario 4, three different numbers of imputations were studied to know how they affected the performances of the two proposed MI estimation methods. Finally, Scenario 5 provided six different correlation coefficients between X_1 and X_2 under the same selection probabilities and sample size to investigate the performances of all the methods in these situations.

For each experimental configuration, 1,000 replications were performed. $M = 30$ was considered except considering $M = 5, 25$, and 45 in Scenario 4. Calculated were the bias, standard deviation (SD), asymptotic standard error (ASE), and coverage probability (CP) of a 95% confidence interval for each estimator. To evaluate the relative efficiencies (REs) of estimators, we computed the ratio of mean square error (MSE) of each of the other estimators, without including the full data ML estimator, to those of the MI1 and MI2 estimators, respectively, where the MSE of an estimator was defined as the sum of the square of bias and the square of SD, i.e., $MSE = \text{bias}^2 + SD^2$.

Scenario 1. The goal was to assess how well all the estimation methods performed when the two covariates X_1 and X_2 were independent by using the same observed selection probabilities and considering the three different sample sizes $n = 500, 1,000$, and 2,000. The discrete distribution of the four values

(−0.3, −0.1, 0.4, 1) with probabilities (0.2, 0.3, 0.3, 0.2), respectively, was used to generate the data of X_1 . The data of X_2 were generated from the discrete distribution of the four values (−1, −0.4, 0.2, 0.6) with probabilities (0.1, 0.3, 0.3, 0.3), respectively. The Bernoulli distribution with success probability 0.4 was used to generate the data of Z . The surrogate variables of X_1 and X_2 were W_k that is 1 if $X_k > 0$; 0 if $X_k \leq 0$, $k = 1, 2$. The Bernoulli distribution with success probability $P(Y = 1|X_1, X_2, Z) = H(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 Z)$ was used to generate the data of Y , where $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^T = (1, -0.5, 1, \log(2))^T$. The following multinomial logistic regression model

$$\ln \left(\frac{P(\delta_{ij} = 1|Y_i, W_{1i}, W_{2i}, Z_i)}{P(\delta_{i4} = 1|Y_i, W_{1i}, W_{2i}, Z_i)} \right) = \alpha_j + \gamma_1 Y_i + \gamma_2 W_{1i} + \gamma_3 W_{2i} + \gamma_4 Z_i \quad (22)$$

was used to generate the data of δ_{ij} given $(Y_i, W_{1i}, W_{2i}, Z_i)$, $i = 1, 2, \dots, n$, $j = 1, 2, 3$, where $\alpha = (\alpha_1, \alpha_2, \alpha_3)^T = (2, 0.6, 0.6)^T$ and $\gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)^T = (0.7, -0.2, 0.1, -1.2)^T$. Under the three different sample sizes, the observed selection probabilities were similar and about 0.6, 0.15, 0.15, and 0.1, respectively. This means that the percentages of complete cases, only X_1 missing, only X_2 missing, and both of them missing were 60%, 15%, 15%, and 10%, respectively.

The simulation results of Scenario 1 are illustrated in Table 1. The full data ML method overall outperformed the other approaches, but it is only considered as a benchmark comparison because it has the practical disadvantage that no missing data are required. The biases of the CC (for β_0, β_3) and MICE (for β_1, β_2) estimators were the largest, which implies that these two methods have the worst estimation. The performances of the SIPW and VL methods were similar, and the MI1, MI2 and JCL methods had similar performances, particularly when $n = 1,000$ and $2,000$. The SD and ASE of each estimator were similar except the MICE estimator for β_1 and β_2 , and decreased when the sample size was increased. The ASEs of the MI1 and MI2 estimators were similar and the smallest compared to the other estimators except the full data ML estimator. The empirical CPs for all the estimation methods were overall close to the nominal probability 95% except the CC (for β_0 when $n = 2,000$; β_3 when $n = 1,000, 2,000$) and MICE (for β_0 when $n = 2,000$; β_1 when $n = 500, 2,000$; β_2 when $n = 1,000, 2,000$) methods. In addition, it can be seen from Table 2 that the relative efficiency values were larger than 1 except the JCL (for β_0 when $n = 500, 1,000$; β_1, β_2 when $n = 500, 1,000, 2,000$) and MICE (for β_0 when $n = 500, 1,000$; β_1 when $n = 500$ and MICE versus MI1; β_3 when $n = 500, 1,000, 2,000$) estimators, which shows that the two proposed MI estimators were comparable with the JCL and MICE estimators (for β_0 as $n = 500, 1,000$; β_1 as $n = 500$; β_3 as $n = 500, 1,000, 2,000$) in terms of efficiency. The relative efficiency values of the MICE estimator to the two proposed MI estimators tended to be increased and were larger than 1 (for β_1, β_2) when the sample size was increased, i.e., the two different types of MI estimators are more efficient than the MICE estimator for β_1 and β_2 when the sample size was increased.

Scenario 2. In this scenario, we examined the impact of observed selection probabilities, i.e., when the missing rates were changed, on the efficiencies of the

Table 1 Simulation results of scenario 1 ($M = 30$; $n = 500, 1,000, 2,000$); $\beta = (1, -0.5, 1, \log(2))^T$; $\alpha = (2, 0.6, 0.6)^T$; $\gamma = (0.7, -0.2, 0.1, -1.2)^T$; observed selection probabilities: (0.6, 0.15, 0.15, 0.1)

n			$\hat{\beta}_F$	$\hat{\beta}_C$	$\hat{\beta}_W$	$\hat{\beta}_V$	$\hat{\beta}_J$	$\hat{\beta}_{M1}$	$\hat{\beta}_{M2}$	$\hat{\beta}_{ME}$
500	β_0	Bias	0.0053	0.0523	0.0093	0.0071	0.0042	0.0077	0.0071	-0.0299
		SD	0.1470	0.1871	0.1496	0.1506	0.1478	0.1480	0.1479	0.1427
		ASE	0.1460	0.1874	0.1492	0.1512	0.1473	0.1470	0.1470	0.1485
		CP	0.9450	0.9520	0.9440	0.9500	0.9410	0.9420	0.9430	0.9520
	β_1	Bias	-0.0069	0.0003	-0.0177	-0.0124	-0.0004	-0.0137	-0.0119	0.0780
		SD	0.2222	0.2931	0.2495	0.2463	0.2293	0.2363	0.2344	0.2224
		ASE	0.2291	0.3020	0.2486	0.2542	0.2370	0.2355	0.2360	0.2602
		CP	0.9660	0.9610	0.9530	0.9620	0.9690	0.9580	0.9570	0.9700
	β_2	Bias	0.0095	0.0138	0.0176	0.0124	-0.0002	0.0131	0.0129	-0.1293
		SD	0.2121	0.2705	0.2318	0.2332	0.2192	0.2227	0.2215	0.2106
		ASE	0.2069	0.2726	0.2217	0.2271	0.2139	0.2116	0.2121	0.2368
		CP	0.9430	0.9550	0.9380	0.9400	0.9400	0.9330	0.9370	0.9410
β_3	Bias	0.0177	0.1223	0.0289	0.0103	0.0180	0.0187	0.0186	-0.0035	
	SD	0.2298	0.3242	0.2411	0.2405	0.2351	0.2317	0.2316	0.2295	
	ASE	0.2321	0.3177	0.2345	0.2459	0.2353	0.2328	0.2329	0.2321	
	CP	0.9560	0.9390	0.9490	0.9580	0.9560	0.9570	0.9600	0.9490	
1,000	β_0	Bias	0.0098	0.0534	0.0091	0.0092	0.0086	0.0095	0.0094	-0.0256
		SD	0.1034	0.1309	0.1063	0.1065	0.1046	0.1053	0.1051	0.1013
		ASE	0.1030	0.1318	0.1051	0.1058	0.1040	0.1038	0.1038	0.1047
		CP	0.9440	0.9430	0.9460	0.9460	0.9410	0.9450	0.9430	0.9470
	β_1	Bias	-0.0027	0.0088	0.0006	0.0004	0.0019	-0.0011	-0.0007	0.0792
		SD	0.1599	0.2133	0.1744	0.1757	0.1656	0.1685	0.1673	0.1643
		ASE	0.1611	0.2112	0.1748	0.1763	0.1666	0.1661	0.1664	0.1828
		CP	0.9460	0.9540	0.9500	0.9480	0.9440	0.9420	0.9440	0.9500
	β_2	Bias	0.0033	0.0047	0.0041	0.0040	-0.0025	0.0031	0.0030	-0.1344
		SD	0.1469	0.1913	0.1580	0.1589	0.1518	0.1542	0.1540	0.1488
		ASE	0.1455	0.1908	0.1561	0.1579	0.1501	0.1493	0.1495	0.1692
		CP	0.9480	0.9480	0.9400	0.9390	0.9390	0.9360	0.9420	0.8920
β_3	Bias	0.0043	0.0925	0.0056	0.0040	0.0064	0.0048	0.0048	-0.0160	
	SD	0.1669	0.2253	0.1684	0.1695	0.1690	0.1678	0.1679	0.1655	
	ASE	0.1632	0.2215	0.1642	0.1678	0.1645	0.1635	0.1635	0.1630	
	CP	0.9480	0.9290	0.9430	0.9480	0.9500	0.9470	0.9450	0.9350	
2,000	β_0	Bias	0.0001	0.0404	0.0006	0.0002	0.0000	0.0003	0.0003	-0.0363
		SD	0.0727	0.0948	0.0745	0.0747	0.0740	0.0737	0.0737	0.0717
		ASE	0.0725	0.0926	0.0740	0.0742	0.0732	0.0731	0.0731	0.0738
		CP	0.9520	0.9200	0.9510	0.9480	0.9470	0.9510	0.9510	0.9250
	β_1	Bias	0.0022	0.0171	0.0003	0.0014	0.0035	0.0015	0.0019	0.0875
		SD	0.1136	0.1472	0.1224	0.1230	0.1174	0.1186	0.1185	0.1142
		ASE	0.1135	0.1485	0.1232	0.1236	0.1176	0.1173	0.1173	0.1292
		CP	0.9550	0.9560	0.9510	0.9540	0.9530	0.9520	0.9500	0.9260

Table 1 (continued)

n		$\hat{\beta}_F$	$\hat{\beta}_C$	$\hat{\beta}_W$	$\hat{\beta}_V$	$\hat{\beta}_J$	$\hat{\beta}_{M1}$	$\hat{\beta}_{M2}$	$\hat{\beta}_{ME}$
β_2	Bias	0.0008	-0.0020	0.0005	0.0003	-0.0013	0.0010	0.0011	-0.1383
	SD	0.1019	0.1389	0.1105	0.1108	0.1050	0.1068	0.1063	0.1054
	ASE	0.1025	0.1340	0.1100	0.1106	0.1057	0.1053	0.1054	0.1224
	CP	0.9490	0.9510	0.9470	0.9420	0.9500	0.9450	0.9430	0.8100
β_3	Bias	0.0044	0.0947	0.0047	0.0045	0.0051	0.0047	0.0046	-0.0157
	SD	0.1150	0.1543	0.1161	0.1165	0.1161	0.1155	0.1154	0.1134
	ASE	0.1149	0.1556	0.1156	0.1167	0.1157	0.1152	0.1152	0.1147
	CP	0.9480	0.9080	0.9470	0.9460	0.9440	0.9470	0.9470	0.9470

Table 2 Relative efficiencies in scenario 1 ($M = 30$; $n = 500, 1,000, 2,000$); $\beta = (1, -0.5, 1, \log(2))^T$; $\alpha = (2, 0.6, 0.6)^T$; $\gamma = (0.7, -0.2, 0.1, -1.2)^T$; observed selection probabilities: (0.6, 0.15, 0.15, 0.1)

n	β	RE_{C1}	RE_{W1}	RE_{V1}	RE_{J1}	RE_{E1}	RE_{C2}	RE_{W2}	RE_{V2}	RE_{J2}	RE_{E2}	RE_{I2}
500	β_0	1.718	1.023	1.035	0.995	0.968	1.722	1.025	1.037	0.997	0.970	1.002
	β_1	1.533	1.117	1.086	0.938	0.992	1.560	1.136	1.104	0.955	1.008	1.017
	β_2	1.474	1.086	1.096	0.966	1.227	1.490	1.098	1.108	0.976	1.241	1.011
	β_3	2.222	1.091	1.072	1.029	0.975	2.224	1.092	1.073	1.030	0.976	1.001
1000	β_0	1.788	1.018	1.022	0.985	0.977	1.795	1.022	1.026	0.989	0.981	1.004
	β_1	1.605	1.071	1.087	0.966	1.171	1.628	1.087	1.103	0.980	1.188	1.014
	β_2	1.539	1.050	1.062	0.969	1.691	1.544	1.053	1.065	0.972	1.695	1.003
	β_3	2.105	1.007	1.020	1.015	0.981	2.103	1.006	1.019	1.014	0.980	0.999
2000	β_0	1.955	1.022	1.027	1.008	1.189	1.955	1.022	1.027	1.008	1.189	1.000
	β_1	1.561	1.065	1.075	0.981	1.471	1.563	1.067	1.077	0.982	1.474	1.002
	β_2	1.692	1.070	1.076	0.967	2.651	1.708	1.081	1.086	0.976	2.676	1.009
	β_3	2.453	1.010	1.017	1.011	0.981	2.457	1.012	1.019	1.013	0.983	1.002

RE_{mk} , $m \in \{C, W, V, J, E\}$, $k = 1, 2$, are the relative efficiencies of the CC, SIPW, VL, JCL, and MICE estimators to MLk estimators for each β_i , $i = 0, 1, 2, 3$.

RE_{mk} is the ratio of MSE of m estimator to MSE of MLk estimator for each β_i .

$$RE_{mk} = MSE_m / MSE_k, RE_{I2} = MSE_1 / MSE_2, MSE = bias^2 + SD^2$$

estimators given $n = 1,000$ fixed. The same as Scenario 1 were the values for β and γ and the procedure to generate the data of X_1, X_2, Z, Y, W_1 , and W_2 . For the multinomial logit model in (22) to generate the data of δ_{ij} , set $\gamma = (0.7, -0.2, 0.1, -1.2)^T$ and $\alpha = (0.7, 0.5, 0.5)^T$, $(1.5, 0.6, 0.6)^T$, and $(2.5, 0.6, 0.6)^T$ to obtain the three sets of observed selection probabilities, (0.31, 0.26, 0.26, 0.17), (0.47, 0.20, 0.20, 0.13), and (0.72, 0.11, 0.11, 0.06), respectively.

Table 3 shows the simulation results of this scenario. The biases, SDs and ASEs of the last seven estimators overall tended to be decreased when the CC percentage was increased from 31% to 72%. The serious bias still happened to the CC (for β_0, β_3) and MICE (for β_1, β_2) estimators. The performances of the MI1 and MI2 estimation

Table 3 Simulation results of scenario 2 ($M = 30$; $n = 1,000$); $\beta = (1, -0.5, 1, \log(2))^T$; $\gamma = (0.7, -0.2, 0.1, -1.2)^T$

		$\hat{\beta}_F$	$\hat{\beta}_C$	$\hat{\beta}_W$	$\hat{\beta}_V$	$\hat{\beta}_J$	$\hat{\beta}_{M1}$	$\hat{\beta}_{M2}$	$\hat{\beta}_{ME}$
Observed selection probabilities: 0.31, 0.26, 0.26, 0.17; $\alpha = (0.7, 0.5, 0.5)^T$									
β_0	Bias	0.0098	0.1027	0.0111	0.0089	0.0069	0.0107	0.0099	-0.0473
	SD	0.1034	0.1910	0.1124	0.1130	0.1069	0.1088	0.1077	0.1013
	ASE	0.1030	0.1835	0.1099	0.1123	0.1050	0.1043	0.1044	0.1060
	CP	0.9440	0.9200	0.9480	0.9480	0.9460	0.9380	0.9420	0.9350
β_1	Bias	-0.0027	0.0094	-0.0029	0.0012	0.0118	-0.0026	-0.0003	0.1315
	SD	0.1599	0.3072	0.2113	0.2047	0.1696	0.1889	0.1827	0.1726
	ASE	0.1611	0.3002	0.2022	0.2092	0.1705	0.1688	0.1695	0.1989
	CP	0.9460	0.9580	0.9410	0.9570	0.9500	0.9160	0.9230	0.9290
β_2	Bias	0.0033	0.0026	0.0147	0.0046	-0.0162	0.0084	0.0061	-0.2332
	SD	0.1469	0.2729	0.1875	0.1842	0.1554	0.1695	0.1633	0.1489
	ASE	0.1455	0.2712	0.1774	0.1848	0.1545	0.1514	0.1519	0.1896
	CP	0.9480	0.9550	0.9330	0.9470	0.9440	0.9170	0.9300	0.8010
β_3	Bias	0.0043	0.1543	0.0231	0.0036	0.0128	0.0057	0.0051	-0.0255
	SD	0.1669	0.3300	0.1911	0.1802	0.1758	0.1696	0.1687	0.1648
	ASE	0.1632	0.3263	0.1671	0.1811	0.1669	0.1638	0.1638	0.1628
	CP	0.9480	0.9390	0.9140	0.9550	0.9310	0.9390	0.9430	0.9330
Observed selection probabilities: 0.47, 0.20, 0.20, 0.13; $\alpha = (1.5, 0.6, 0.6)^T$									
β_0	Bias	0.0098	0.0759	0.0098	0.0094	0.0077	0.0095	0.0091	-0.0361
	SD	0.1034	0.1495	0.1072	0.1076	0.1043	0.1054	0.1054	0.1024
	ASE	0.1030	0.1478	0.1065	0.1076	0.1043	0.1040	0.1040	0.1053
	CP	0.9440	0.9240	0.9480	0.9510	0.9400	0.9460	0.9450	0.9440
β_1	Bias	-0.0027	0.0095	0.0003	0.0012	0.0071	0.0001	0.0011	0.1089
	SD	0.1599	0.2345	0.1839	0.1841	0.1668	0.1737	0.1723	0.1662
	ASE	0.1611	0.2384	0.1833	0.1859	0.1683	0.1674	0.1678	0.1896
	CP	0.9460	0.9560	0.9520	0.9500	0.9470	0.9400	0.9500	0.9440
β_2	Bias	0.0033	0.0106	0.0075	0.0067	-0.0054	0.0061	0.0045	-0.1731
	SD	0.1469	0.2179	0.1670	0.1677	0.1533	0.1576	0.1562	0.1502
	ASE	0.1455	0.2150	0.1623	0.1653	0.1515	0.1503	0.1506	0.1763
	CP	0.9480	0.9550	0.9420	0.9350	0.9370	0.9350	0.9360	0.8600
β_3	Bias	0.0043	0.1159	0.0071	0.0055	0.0092	0.0057	0.0053	-0.0189
	SD	0.1669	0.2694	0.1703	0.1711	0.1698	0.1678	0.1676	0.1645
	ASE	0.1632	0.2530	0.1649	0.1711	0.1653	0.1636	0.1636	0.1629
	CP	0.9480	0.9080	0.9390	0.9530	0.9410	0.9410	0.9400	0.9400
Observed selection probabilities: 0.72, 0.11, 0.11, 0.06; $\alpha = (2.5, 0.6, 0.6)^T$									
β_0	Bias	0.0098	0.0405	0.0106	0.0106	0.0093	0.0102	0.0101	-0.0175
	SD	0.1034	0.1234	0.1049	0.1048	0.1039	0.1043	0.1043	0.1016
	ASE	0.1030	0.1212	0.1044	0.1048	0.1037	0.1036	0.1036	0.1043
	CP	0.9440	0.9320	0.9480	0.9470	0.9410	0.9460	0.9450	0.9480

Table 3 (continued)

		$\hat{\beta}_F$	$\hat{\beta}_C$	$\hat{\beta}_W$	$\hat{\beta}_V$	$\hat{\beta}_J$	$\hat{\beta}_{M1}$	$\hat{\beta}_{M2}$	$\hat{\beta}_{ME}$
β_1	Bias	-0.0027	0.0071	-0.0030	-0.0033	-0.0001	-0.0031	-0.0025	0.0604
	SD	0.1599	0.1931	0.1705	0.1701	0.1643	0.1671	0.1662	0.1583
	ASE	0.1611	0.1929	0.1696	0.1705	0.1652	0.1649	0.1651	0.1772
	CP	0.9460	0.9530	0.9520	0.9510	0.9470	0.9460	0.9480	0.9630
β_2	Bias	0.0033	0.0043	0.0063	0.0054	0.0010	0.0052	0.0052	-0.0987
	SD	0.1469	0.1726	0.1535	0.1541	0.1492	0.1501	0.1497	0.1448
	ASE	0.1455	0.1742	0.1520	0.1530	0.1488	0.1484	0.1485	0.1618
	CP	0.9480	0.9530	0.9410	0.9400	0.9470	0.9460	0.9460	0.9240
β_3	Bias	0.0043	0.0741	0.0048	0.0047	0.0055	0.0049	0.0050	-0.0108
	SD	0.1669	0.2084	0.1674	0.1685	0.1680	0.1670	0.1669	0.1653
	ASE	0.1632	0.2004	0.1638	0.1659	0.1640	0.1634	0.1634	0.1631
	CP	0.9480	0.9330	0.9440	0.9470	0.9450	0.9450	0.9460	0.9440

methods were essentially the same, and the ASEs of the two proposed MI estimators were overall the smallest in comparison with the other estimators except the full data ML estimator. The empirical CPs based on all the estimation methods were overall close to the nominal probability 95% except the CC (for β_0 when CC percentage = 0.31, 0.47), SIPW (for β_3 when CC percentage = 0.31), MI1 and MI2 (for β_1, β_2 when CC percentage = 0.31), and MICE (for β_1 when CC percentage = 0.31; β_2 when CC percentage = 0.31, 0.47, 0.72) methods. The relative efficiency values in Table 4 were still greater than 1 except the JCL (for $\beta_0, \beta_1, \beta_2$) and MICE (for β_0 when CC percentage = 0.72; β_3) estimators, and tended to decrease except the JCL (for $\beta_0, \beta_1, \beta_2$) and MICE (for β_3) estimators when the CC percentage was increased, indicating that the two different types of MI estimators were comparable with the JCL estimator and the most efficient compared to the CC, SIPW, VL, and MICE estimators. When the CC percentage was 0.72, the relative efficiency values were (very) close to 1 for the SIPW, VL, and JCL estimators compared to the two MI estimators.

Scenario 3. In this scenario, we wished to know whether changing the values of the logistic regression parameters affects the performances of the estimation methods. Therefore, we kept all the settings in Scenario 2 except changing $\beta = (1, -0.5, 1, \log(2))^T$ to $\beta = (-1, 1, 0.7, -1)^T$. As the simulation results given in Tables 5 and 6, the three sets of observed selection probabilities were (0.29, 0.24, 0.24, 0.23), (0.45, 0.19, 0.19, 0.16), and (0.70, 0.10, 0.10, 0.10), respectively, which were quite similar to those in Scenario 2, although the first three observed selection probabilities 0.29, 0.24, and 0.24 in the first set in this scenario were slightly reduced, compared to the first three observed selection probabilities 0.31, 0.26, and 0.26 in the first set in Scenario 2, and the last observed selection probability 0.23 was slightly increased compared to the last one 0.17. The performances of all these estimation methods in this scenario were quite similar to those in Scenario 2. One of the reasons might have the similar missing rates in this scenario

Table 4 Relative efficiencies in scenario 2 ($M = 30$; $n = 1,000$); $\beta = (1, -0.5, 1, \log(2))^T$; $\gamma = (0.7, -0.2, 0.1, -1.2)^T$

β	RE_{C1}	RE_{W1}	RE_{V1}	RE_{J1}	RE_{E1}	RE_{C2}	RE_{W2}	RE_{V2}	RE_{J2}	RE_{E2}	RE_{12}
Observed selection probabilities: 0.31, 0.26, 0.26, 0.17; $\alpha = (0.7, 0.5, 0.5)^T$											
β_0	3.935	1.067	1.075	0.960	1.046	4.020	1.091	1.099	0.981	1.069	1.022
β_1	2.647	1.251	1.174	0.810	1.319	2.830	1.338	1.255	0.866	1.410	1.069
β_2	2.586	1.228	1.179	0.848	2.658	2.789	1.325	1.271	0.914	2.866	1.079
β_3	4.608	1.287	1.128	1.079	0.966	4.659	1.301	1.140	1.091	0.976	1.011
Observed selection probabilities: 0.47, 0.20, 0.20, 0.13; $\alpha = (1.5, 0.6, 0.6)^T$											
β_0	2.510	1.035	1.042	0.977	1.053	2.511	1.035	1.042	0.977	1.054	1.001
β_1	1.826	1.121	1.123	0.924	1.309	1.855	1.139	1.142	0.939	1.330	1.016
β_2	1.913	1.123	1.132	0.946	2.112	1.949	1.144	1.153	0.964	2.152	1.019
β_3	3.051	1.031	1.040	1.026	0.973	3.059	1.033	1.042	1.028	0.975	1.002
Observed selection probabilities: 0.72, 0.11, 0.11, 0.06; $\alpha = (2.5, 0.6, 0.6)^T$											
β_0	1.536	1.012	1.010	0.991	0.968	1.536	1.012	1.010	0.991	0.968	1.000
β_1	1.337	1.041	1.036	0.966	1.028	1.351	1.052	1.048	0.977	1.039	1.011
β_2	1.321	1.046	1.054	0.987	1.361	1.329	1.052	1.060	0.992	1.368	1.005
β_3	1.752	1.005	1.018	1.012	0.983	1.754	1.006	1.019	1.013	0.984	1.001

RE_{mk} , $m \in \{C, W, V, J, E\}$, $k = 1, 2$, are the relative efficiencies of the CC, SIPW, VL, JCL, and MICE estimators to MIk estimators for each β_i , $i = 0, 1, 2, 3$.

RE_{mk} is the ratio of MSE of m estimator to MSE of MIk estimator for each β_i .

$$RE_{mk} = MSE_m / MSE_k, RE_{12} = MSE_1 / MSE_2, MSE = bias^2 + SD^2$$

and Scenario 2. Hence, in general, changing the value of β had a little effect on the selection probabilities and the efficiencies of the proposed estimation methods.

Scenario 4. The purpose of this scenario was to examine the impact of the number of multiple imputations on the performances of the proposed methods when X_1 and X_2 were independent with a fixed sample size. All the settings were the same as in Scenario 1 except considering $M = 5, 25$, and 45 , respectively, given $n = 1,000$. The observed selection probabilities were 0.60, 0.15, 0.15, and 0.10.

The simulation results for the MI1, MI2 and MICE methods in Table 7 were essentially the same as those in Table 1 of Scenario 1 when $n = 1,000$. Table 8 shows that the relative efficiency values were also overall the same as those in Table 2 of Scenario 1 for $n = 1,000$. Therefore, this simulation study demonstrated that the two proposed MI methods were not affected by the number of imputations.

Table 9 provides a summary of computing time for each estimation method. The JCL and MICE methods had the longest and second longest computing time, respectively. The JCL method took an average of 21.88 seconds to perform one simulation, which are approximately 73, 35.3 and 23 times the two proposed MI methods when $M = 5, 25$, and 45 , respectively. On average, the MICE methods took 1.15, 5.73, and 10.30 seconds to perform one simulation when $M = 5, 25$, and 45 , which are approximately 3.8, 9.2 and 10.8 times the two proposed MI methods, respectively. However, the performances of the JCL and two proposed MI methods were essentially the same. Therefore, based on the simulation results, the MI1 and MI2

Table 5 Simulation results of scenario 3 ($M = 30$; $n = 1,000$); $\beta = (-1, 1, 0.7, -1)^T$; $\gamma = (0.7, -0.2, 0.1, -1.2)^T$

		$\hat{\beta}_F$	$\hat{\beta}_C$	$\hat{\beta}_W$	$\hat{\beta}_V$	$\hat{\beta}_J$	$\hat{\beta}_{M1}$	$\hat{\beta}_{M2}$	$\hat{\beta}_{ME}$
Observed selection probabilities: 0.29, 0.24, 0.24, 0.23; $\alpha = (0.7, 0.5, 0.5)^T$									
β_0	Bias	-0.0067	0.0628	-0.0133	-0.0069	-0.0032	-0.0092	-0.0073	0.0617
	SD	0.1022	0.1810	0.1116	0.1110	0.1040	0.1065	0.1049	0.0990
	ASE	0.1029	0.1802	0.1109	0.1130	0.1048	0.1044	0.1045	0.1075
	CP	0.9520	0.9240	0.9560	0.9600	0.9570	0.9600	0.9530	0.9240
β_1	Bias	0.0088	0.0411	0.0231	0.0072	0.0001	0.0128	0.0079	-0.1815
	SD	0.1599	0.2935	0.2106	0.2035	0.1704	0.1864	0.1797	0.1715
	ASE	0.1608	0.2906	0.2038	0.2074	0.1719	0.1689	0.1695	0.2023
	CP	0.9620	0.9530	0.9420	0.9550	0.9580	0.9230	0.9400	0.8800
β_2	Bias	0.0111	0.0169	0.0296	0.0082	0.0062	0.0217	0.0156	-0.1490
	SD	0.1577	0.2801	0.1971	0.1917	0.1668	0.1790	0.1740	0.1577
	ASE	0.1557	0.2809	0.1877	0.1953	0.1646	0.1618	0.1625	0.1905
	CP	0.9560	0.9580	0.9430	0.9530	0.9600	0.9340	0.9450	0.9100
β_3	Bias	-0.0148	0.1034	-0.0428	-0.0088	-0.0112	-0.0162	-0.0152	0.0191
	SD	0.1701	0.3312	0.1891	0.1784	0.1715	0.1713	0.1708	0.1661
	ASE	0.1676	0.3179	0.1729	0.1852	0.1705	0.1683	0.1684	0.1677
	CP	0.9430	0.9300	0.9300	0.9570	0.9460	0.9430	0.9420	0.9470
Observed selection probabilities: 0.46, 0.19, 0.19, 0.16; $\alpha = (1.5, 0.6, 0.6)^T$									
β_0	Bias	-0.0067	0.0443	-0.0102	-0.0074	-0.0055	-0.0089	-0.0081	0.0426
	SD	0.1022	0.1435	0.1068	0.1072	0.1043	0.1049	0.1048	0.1021
	ASE	0.1029	0.1454	0.1069	0.1079	0.1042	0.1041	0.1041	0.1062
	CP	0.9520	0.9430	0.9490	0.9550	0.9560	0.9490	0.9500	0.9410
β_1	Bias	0.0088	0.0314	0.0196	0.0130	0.0066	0.0154	0.0131	-0.1290
	SD	0.1599	0.2335	0.1862	0.1854	0.1704	0.1748	0.1736	0.1686
	ASE	0.1608	0.2324	0.1834	0.1853	0.1689	0.1676	0.1679	0.1906
	CP	0.9620	0.9450	0.9620	0.9660	0.9520	0.9440	0.9450	0.9040
β_2	Bias	0.0111	0.0141	0.0127	0.0051	0.0073	0.0104	0.0096	-0.1095
	SD	0.1577	0.2290	0.1768	0.1771	0.1634	0.1684	0.1672	0.1604
	ASE	0.1557	0.2246	0.1721	0.1757	0.1618	0.1603	0.1608	0.1810
	CP	0.9560	0.9510	0.9490	0.9520	0.9550	0.9480	0.9560	0.9230
β_3	Bias	-0.0148	0.0748	-0.0216	-0.0123	-0.0122	-0.0153	-0.0152	0.0123
	SD	0.1701	0.2576	0.1751	0.1750	0.1708	0.1710	0.1710	0.1669
	ASE	0.1676	0.2499	0.1698	0.1759	0.1691	0.1682	0.1682	0.1679
	CP	0.9430	0.9290	0.9440	0.9550	0.9470	0.9460	0.9450	0.9440
Observed selection probabilities: 0.70, 0.10, 0.10, 0.10; $\alpha = (2.5, 0.6, 0.6)^T$									
β_0	Bias	-0.0067	0.0217	-0.0076	-0.0068	-0.0061	-0.0071	-0.0071	0.0217
	SD	0.1022	0.1184	0.1038	0.1039	0.1032	0.1031	0.1032	0.1015
	ASE	0.1029	0.1200	0.1043	0.1047	0.1036	0.1035	0.1035	0.1045
	CP	0.9520	0.9490	0.9510	0.9500	0.9520	0.9520	0.9530	0.9450

Table 5 (continued)

		$\hat{\beta}_F$	$\hat{\beta}_C$	$\hat{\beta}_W$	$\hat{\beta}_V$	$\hat{\beta}_J$	$\hat{\beta}_{M1}$	$\hat{\beta}_{M2}$	$\hat{\beta}_{ME}$
β_1	Bias	0.0088	0.0199	0.0114	0.0097	0.0083	0.0100	0.0104	-0.0673
	SD	0.1599	0.1904	0.1697	0.1693	0.1645	0.1652	0.1652	0.1606
	ASE	0.1608	0.1899	0.1694	0.1702	0.1651	0.1646	0.1648	0.1761
	CP	0.9620	0.9480	0.9470	0.9500	0.9510	0.9450	0.9460	0.9580
β_2	Bias	0.0111	0.0077	0.0123	0.0115	0.0099	0.0114	0.0113	-0.0580
	SD	0.1577	0.1898	0.1642	0.1643	0.1604	0.1622	0.1619	0.1575
	ASE	0.1557	0.1839	0.1622	0.1635	0.1590	0.1586	0.1587	0.1696
	CP	0.9560	0.9530	0.9520	0.9510	0.9560	0.9490	0.9560	0.9550
β_3	Bias	-0.0148	0.0462	-0.0160	-0.0142	-0.0134	-0.0151	-0.0152	0.0017
	SD	0.1701	0.2045	0.1708	0.1719	0.1703	0.1702	0.1704	0.1680
	ASE	0.1676	0.2014	0.1684	0.1705	0.1683	0.1679	0.1680	0.1679
	CP	0.9430	0.9400	0.9460	0.9470	0.9470	0.9440	0.9460	0.9480

Table 6 Relative efficiencies in scenario 3 ($M = 30$; $n = 1,000$); $\beta = (-1, 1, 0.7, -1)^T$; $\gamma = (0.7, -0.2, 0.1, -1.2)^T$

β	RE_{C1}	RE_{W1}	RE_{V1}	RE_{J1}	RE_{E1}	RE_{C2}	RE_{W2}	RE_{V2}	RE_{J2}	RE_{E2}	RE_{12}
Observed selection probabilities: 0.29, 0.24, 0.24, 0.23; $\alpha = (0.7, 0.5, 0.5)^T$											
β_0	3.212	1.105	1.082	0.947	1.190	3.319	1.142	1.119	0.979	1.230	1.033
β_1	2.516	1.286	1.188	0.832	1.786	2.715	1.387	1.282	0.897	1.927	1.079
β_2	2.422	1.222	1.133	0.857	1.448	2.580	1.302	1.206	0.913	1.543	1.065
β_3	4.066	1.270	1.078	0.998	0.944	4.094	1.278	1.085	1.004	0.951	1.007
Observed selection probabilities: 0.46, 0.19, 0.19, 0.16; $\alpha = (1.5, 0.6, 0.6)^T$											
β_0	2.035	1.038	1.042	0.984	1.104	2.042	1.042	1.045	0.987	1.108	1.003
β_1	1.803	1.138	1.122	0.944	1.464	1.832	1.157	1.140	0.960	1.487	1.016
β_2	1.849	1.104	1.103	0.940	1.325	1.877	1.120	1.119	0.954	1.345	1.015
β_3	2.441	1.056	1.044	0.995	0.950	2.442	1.056	1.044	0.995	0.950	1.000
Observed selection probabilities: 0.70, 0.10, 0.10, 0.10; $\alpha = (2.5, 0.6, 0.6)^T$											
β_0	1.357	1.014	1.015	1.001	1.009	1.354	1.012	1.013	1.000	1.007	0.998
β_1	1.338	1.056	1.050	0.990	1.107	1.338	1.056	1.050	0.990	1.107	1.000
β_2	1.365	1.026	1.026	0.977	1.065	1.370	1.029	1.030	0.981	1.070	1.004
β_3	1.505	1.008	1.019	1.000	0.967	1.502	1.006	1.016	0.997	0.964	0.998

RE_{mk} , $m \in \{C, W, V, J, E\}$, $k = 1, 2$, are the relative efficiencies of the CC, SIPW, VL, JCL, and MICE estimators to MIK estimators for each β_i , $i = 0, 1, 2, 3$.

RE_{mk} is the ratio of MSE of m estimator to MSE of MIK estimator for each β_i .

$$RE_{mk} = MSE_m / MSE_k, RE_{12} = MSE_1 / MSE_2, MSE = bias^2 + SD^2$$

methods overall outperformed the other methods, except the full data ML and JCL methods, when estimating the parameters of logistic regression with covariates MAR.

Table 7 Simulation results of scenario 4 ($n = 1, 000$); $\beta = (1, -0.5, 1, \log(2))^T$; $\alpha = (2, 0.6, 0.6)^T$; $\gamma = (0.7, -0.2, 0.1, -1.2)^T$; observed selection probabilities: $(0.60, 0.15, 0.15, 0.10)$; $M = 5, 25, 45$

		$M = 5$			$M = 25$			$M = 45$				
		$\hat{\beta}_F$	$\hat{\beta}_C$	$\hat{\beta}_W$	$\hat{\beta}_V$	$\hat{\beta}_J$	$\hat{\beta}_{M1}$	$\hat{\beta}_{M2}$	$\hat{\beta}_{ME}$	$\hat{\beta}_{M1}$	$\hat{\beta}_{M2}$	$\hat{\beta}_{ME}$
β_0	Bias	0.0098	0.0534	0.0091	0.0092	0.0086	0.0094	0.0095	-0.0252	0.0095	0.0095	-0.0260
	SD	0.1034	0.1309	0.1063	0.1065	0.1046	0.1052	0.1050	0.1015	0.1053	0.1051	0.1010
	ASE	0.1030	0.1318	0.1051	0.1058	0.1040	0.1039	0.1040	0.1053	0.1038	0.1038	0.1048
	CP	0.9440	0.9430	0.9460	0.9460	0.9410	0.9460	0.9450	0.9440	0.9450	0.9450	0.9440
β_1	Bias	-0.0027	0.0088	0.0006	0.0004	0.0019	-0.0012	-0.0009	0.0779	-0.0012	-0.0010	0.0801
	SD	0.1599	0.2133	0.1744	0.1757	0.1656	0.1686	0.1682	0.1689	0.1685	0.1672	0.1637
	ASE	0.1611	0.2112	0.1748	0.1763	0.1666	0.1670	0.1673	0.1856	0.1662	0.1664	0.1832
	CP	0.9460	0.9540	0.9500	0.9480	0.9440	0.9490	0.9440	0.9440	0.9450	0.9440	0.9470
β_2	Bias	0.0033	0.0047	0.0041	0.0040	-0.0025	0.0035	0.0026	-0.1350	0.0024	0.0033	-0.1353
	SD	0.1469	0.1913	0.1580	0.1589	0.1518	0.1548	0.1543	0.1499	0.1544	0.1534	0.1490
	ASE	0.1455	0.1908	0.1561	0.1579	0.1501	0.1500	0.1500	0.1717	0.1494	0.1495	0.1692
	CP	0.9480	0.9480	0.9400	0.9390	0.9390	0.9400	0.9380	0.9000	0.9370	0.9410	0.8860
β_3	Bias	0.0043	0.0925	0.0056	0.0040	0.0064	0.0049	0.0047	-0.0161	0.0047	0.0047	-0.0160
	SD	0.1669	0.2253	0.1684	0.1695	0.1690	0.1678	0.1677	0.1656	0.1677	0.1677	0.1656
	ASE	0.1632	0.2215	0.1642	0.1678	0.1645	0.1636	0.1636	0.1633	0.1635	0.1635	0.1631
	CP	0.9480	0.9290	0.9430	0.9480	0.9500	0.9460	0.9460	0.9390	0.9450	0.9460	0.9360

Table 8 Relative efficiencies in scenario 4 ($M = 5, 25, 45$; $n = 1,000$); $\beta = (1, -0.5, 1, \log(2))^T$; $\alpha = (2, 0.6, 0.6)^T$; $\gamma = (0.7, -0.2, 0.1, -1.2)^T$; observed selection probabilities: (0.60, 0.15, 0.15, 0.10)

M	β	RE_{C1}	RE_{W1}	RE_{V1}	RE_{J1}	RE_{E1}	RE_{C2}	RE_{W2}	RE_{V2}	RE_{J2}	RE_{E2}	RE_{12}
5	β_0	1.792	1.020	1.024	0.987	0.981	1.798	1.024	1.028	0.991	0.984	1.004
	β_1	1.603	1.070	1.086	0.965	1.217	1.611	1.075	1.091	0.969	1.223	1.005
	β_2	1.527	1.042	1.054	0.961	1.697	1.538	1.049	1.061	0.968	1.709	1.007
	β_3	2.105	1.007	1.020	1.015	0.982	2.108	1.009	1.021	1.016	0.984	1.001
25	β_0	1.788	1.018	1.022	0.986	0.973	1.795	1.022	1.026	0.989	0.977	1.004
	β_1	1.605	1.071	1.087	0.966	1.170	1.630	1.088	1.104	0.981	1.188	1.005
	β_2	1.536	1.048	1.060	0.967	1.698	1.555	1.061	1.073	0.979	1.720	1.007
	β_3	2.108	1.009	1.021	1.016	0.983	2.108	1.009	1.021	1.016	0.983	1.001
45	β_0	1.795	1.022	1.026	0.989	0.976	1.798	1.024	1.028	0.991	0.978	1.004
	β_1	1.609	1.074	1.090	0.968	1.153	1.628	1.087	1.103	0.980	1.167	1.005
	β_2	1.540	1.050	1.062	0.969	1.702	1.551	1.058	1.070	0.977	1.715	1.007
	β_3	2.105	1.007	1.020	1.015	0.979	2.105	1.008	1.020	1.015	0.979	1.001

RE_{mk} , $m \in \{C, W, V, J, E\}$, $k = 1, 2$, are the relative efficiencies of the CC, SIPW, VL, JCL, and MICE estimators to MLk estimators for each β_i , $i = 0, 1, 2, 3$.

RE_{mk} is the ratio of MSE of m estimator to MSE of MLk estimator for each β_i .

$$RE_{mk} = MSE_m / MSE_k, RE_{12} = MSE_1 / MSE_2, MSE = bias^2 + SD^2$$

Table 9 Summary of computing time (second) for performing one simulation for each estimation method in Scenario 4 ($M = 5, 25, 45$; $n = 1,000$); $\beta = (1, -0.5, 1, \log(2))^T$; $\alpha = (2, 0.6, 0.6)^T$; $\gamma = (0.7, -0.2, 0.1, -1.2)^T$; observed selection probabilities: (0.60, 0.15, 0.15, 0.10)

	$\hat{\beta}_F$	$\hat{\beta}_C$	$\hat{\beta}_W$	$\hat{\beta}_V$	$\hat{\beta}_J$	$M = 5$			$M = 25$			$M = 45$		
						$\hat{\beta}_{M1}$	$\hat{\beta}_{M2}$	$\hat{\beta}_{ME}$	$\hat{\beta}_{M1}$	$\hat{\beta}_{M2}$	$\hat{\beta}_{ME}$	$\hat{\beta}_{M1}$	$\hat{\beta}_{M2}$	$\hat{\beta}_{ME}$
Min.	0.01	0.01	0.13	0.15	17.97	0.28	0.29	1.07	0.57	0.57	5.39	0.86	0.89	9.69
Mean	0.02	0.02	0.13	0.16	21.88	0.30	0.30	1.15	0.62	0.63	5.73	0.95	0.96	10.30
Max.	0.06	0.06	0.18	0.20	24.24	0.38	0.36	1.34	0.69	0.69	6.34	1.05	1.09	11.33

Scenario 5. The scenario aimed to examine the performances of all the approaches when X_1 and X_2 were correlated given that the observed selection probabilities and sample size were fixed. The settings of M , β , and γ in this scenario were the same as those in Scenario 1. We set $\alpha = (1.15, 0.6, 0.6)^T$ to compare the efficiencies of all the estimation methods in a situation of higher missing rates, i.e., the observed selection probabilities (0.40, 0.23, 0.23, 0.14), given $n = 1,000$. In addition, to generate the correlated data of X_1 and X_2 , the distribution that had the four values $-0.3, -0.1, 0.4$, and 1 with probabilities $0.2, 0.3, 0.3$, and 0.2 , respectively, was used to first generate the data of X_1 . Then, given each value of X_1 , the data of X_2 , $(-1, -0.4, 0.2, 0.6)$, were generated such that the correlations between X_1 and X_2 were $\rho = -0.21, -0.53, -0.71, 0.21, 0.51$, and 0.71 , respectively. For example, for the case of $\rho = -0.21$, if $X_1 = -0.3$, the four values of X_2 were generated with

probabilities 0.42, 0.1, 0.2, and 0.28, respectively. Similarly, when $X_1 = -0.1, 0.4$, and 1, we generated the four values of X_2 with probabilities (0.2, 0.1, 0.4, 0.3), (0.2, 0.4, 0.2, 0.2), and (0.65, 0.05, 0.05, 0.25), respectively. Finally, the surrogate variables W_1 and W_2 of X_1 and X_2 , respectively, were set as in Scenario 1.

Given in Tables 10 and 11, respectively, are the simulation results for X_1 and X_2 with negative and positive correlation values. The biases of the CC and MICE estimators were overall larger than the other estimators. The biases were overall not changed much when the absolute correlation value was increased except the MICE estimator (for β_1 when ρ was changed from -0.21 to -0.71 ; $\beta_0, \beta_1, \beta_2$ when ρ was changed from 0.21 to 0.71). The SDs and ASEs of all the estimators, except the estimators of β_3 , tended to be increased with the increase in the absolute correlation value. The SD and ASE of the CC estimator were the largest. The SD and ASE of the JCL estimator were comparable with or (slightly) larger than those of the MI1 and MI2 estimators. The empirical CPs based on all the estimation methods were overall close to the nominal probability 95% except the CC (for β_0 when $\rho = -0.53$; β_3 when $\rho = -0.53, -0.71, 0.51, 0.71$), MI1 (for β_0 when $\rho = 0.71$; β_1, β_2 when $\rho = -0.71, 0.51$), MI2 (for β_1 when $\rho = -0.71, 0.51$), and MICE (for β_0 when $\rho = -0.21, 0.21, 0.51, 0.71$; β_1, β_2 for all ρ values) methods. The relative efficiency values of the CC, SIPW, VL, and MICE estimators to the two different types of MI estimators were greater than 1 except the relative efficiency values of the SIPW to MI2 estimators and the MICE to the two different types of MI estimators (for β_0 when $\rho = -0.53, -0.71$; β_1 when $\rho = -0.21$; β_3 when $\rho = 0.21, 0.51, 0.71$). The performances of the MI1 and MI2 estimators were comparable with that of the JCL estimator in terms of efficiency (Table 12).

In summary, Scenarios 1–5 show that the performances of the JCL and two proposed MI methods were comparable, but Scenario 4 demonstrates that the MI1 and MI2 methods shortened the computing time the JCL and MICE methods took.

5.2 Examples

Two real data examples were used to demonstrate the practicality of the two proposed MI methods and CC, SIPW, VL, JCL, and MICE methods. $M = 30$ imputations were used.

5.2.1 Example 1

The first real data example is the data set of the Global Longitudinal Study of Osteoporosis in Women (GLOW500M) (Hosmer et al. 2013). Let Y be a binary outcome variable to denote whether a respondent had any fracture in the first year, where $Y = 1$ if yes; $Y = 0$ if no. Three covariates are considered. X_1 is used to indicate the history of prior fracture that had missing values, in which X_1 is 1 if yes and 0 if no. X_2 denotes the self-reported risk of fracture that had missing values, where $X_2 = 1$ if less than others of the same age; 2 if same as others of the same age; 3 if greater than others of the same age. Z denotes age at enrollment with integer values from 55 to 90 without missing values. The sample size is $n = 500$. The rates of only X_1 missing, only X_2 missing, and both

Table 10 Simulation results of scenario 5 ($M = 30$; $\rho = -0.21, -0.53, -0.71$; $n = 1,000$); $\beta = (1, -0.5, 1, \log(2))^T$; $\alpha = (1.15, 0.6, 0.6)^T$; $\gamma = (0.7, -0.2, 0.1, -1.2)^T$; observed selection probabilities: (0.40, 0.23, 0.23, 0.14)

ρ			$\hat{\beta}_F$	$\hat{\beta}_C$	$\hat{\beta}_W$	$\hat{\beta}_V$	$\hat{\beta}_J$	$\hat{\beta}_{M1}$	$\hat{\beta}_{M2}$	$\hat{\beta}_{ME}$
-0.21	β_0	Bias	0.0102	0.0830	0.0132	0.0104	0.0094	0.0120	0.0235	-0.0608
		SD	0.1085	0.1672	0.1127	0.1141	0.1120	0.1102	0.1114	0.1073
		ASE	0.1051	0.1668	0.1112	0.1128	0.1069	0.1064	0.1077	0.1114
		CP	0.9470	0.9300	0.9440	0.9490	0.9410	0.9420	0.9400	0.9140
	β_1	Bias	0.0015	0.0171	0.0008	-0.0003	0.0046	0.0020	- 0.0318	-0.0036
		SD	0.1579	0.2544	0.1880	0.1872	0.1649	0.1738	0.1725	0.1626
		ASE	0.1572	0.2579	0.1852	0.1893	0.1643	0.1632	0.1634	0.1920
		CP	0.9570	0.9600	0.9480	0.9550	0.9470	0.9350	0.9350	0.9800
	β_2	Bias	0.0061	0.0080	0.0139	0.0065	0.0017	0.0112	0.0173	- 0.1720
		SD	0.1178	0.1928	0.1356	0.1356	0.1237	0.1274	0.1261	0.1197
		ASE	0.1195	0.1954	0.1345	0.1389	0.1243	0.1227	0.1233	0.1556
		CP	0.9470	0.9530	0.9500	0.9560	0.9470	0.9390	0.9410	0.8400
β_3	Bias	0.0016	0.1324	0.0108	0.0011	0.0095	0.0029	0.0036	-0.0400	
	SD	0.1556	0.2650	0.1654	0.1646	0.1615	0.1576	0.1581	0.1546	
	ASE	0.1560	0.2683	0.1594	0.1668	0.1589	0.1565	0.1566	0.1560	
	CP	0.9560	0.9360	0.9360	0.9480	0.9460	0.9440	0.9450	0.9470	
-0.53	β_0	Bias	0.0077	0.0772	0.0084	0.0081	0.0082	0.0082	0.0150	0.0054
		SD	0.1094	0.1756	0.1124	0.1129	0.1101	0.1103	0.1108	0.1074
		ASE	0.1056	0.1670	0.1090	0.1106	0.1068	0.1063	0.1066	0.1115
		CP	0.9460	0.9230	0.9410	0.9480	0.9440	0.9430	0.9390	0.9590
	β_1	Bias	-0.0013	0.0200	0.0031	0.0056	-0.0007	-0.0004	-0.0280	- 0.2016
		SD	0.1792	0.3050	0.2145	0.2116	0.1884	0.1996	0.1938	0.1686
		ASE	0.1847	0.3023	0.2096	0.2154	0.1924	0.1908	0.1892	0.2237
		CP	0.9600	0.9460	0.9550	0.9550	0.9490	0.9410	0.9480	0.9290
	β_2	Bias	0.0103	0.0092	0.0160	0.0158	0.0051	0.0123	0.0084	- 0.1535
		SD	0.1314	0.2154	0.1499	0.1514	0.1357	0.1418	0.1359	0.1203
		ASE	0.1339	0.2193	0.1492	0.1546	0.1393	0.1376	0.1363	0.1661
		CP	0.9510	0.9550	0.9440	0.9560	0.9540	0.9410	0.9500	0.9160
β_3	Bias	0.0064	0.1362	0.0110	0.0031	0.0062	0.0066	0.0071	-0.0438	
	SD	0.1637	0.2845	0.1736	0.1736	0.1699	0.1660	0.1660	0.1629	
	ASE	0.1638	0.2820	0.1663	0.1745	0.1660	0.1642	0.1644	0.1648	
	CP	0.9500	0.9290	0.9420	0.9440	0.9450	0.9430	0.9460	0.9390	

Table 10 (continued)

ρ			$\hat{\beta}_F$	$\hat{\beta}_C$	$\hat{\beta}_W$	$\hat{\beta}_V$	$\hat{\beta}_J$	$\hat{\beta}_{M1}$	$\hat{\beta}_{M2}$	$\hat{\beta}_{ME}$
-0.71	β_0	Bias	0.0059	0.0793	0.0124	0.0052	0.0043	0.0075	0.0075	-0.0101
		SD	0.1110	0.1683	0.1173	0.1157	0.1119	0.1132	0.1125	0.1096
		ASE	0.1075	0.1705	0.1113	0.1129	0.1086	0.1084	0.1083	0.1177
		CP	0.9350	0.9320	0.9420	0.9420	0.9390	0.9380	0.9380	0.9580
	β_1	Bias	0.0006	0.0009	0.0175	-0.0078	-0.0108	0.0105	0.0049	-0.3282
		SD	0.2065	0.3491	0.3054	0.3034	0.2406	0.2634	0.2532	0.1666
		ASE	0.2134	0.3501	0.2948	0.3052	0.2464	0.2342	0.2343	0.2420
		CP	0.9610	0.9470	0.9420	0.9490	0.9480	0.9180	0.9250	0.8310
	β_2	Bias	0.0095	0.0068	0.0319	0.0041	-0.0030	0.0193	0.0159	-0.1863
		SD	0.1486	0.2518	0.1967	0.1940	0.1646	0.1754	0.1698	0.1092
		ASE	0.1517	0.2483	0.1896	0.1994	0.1691	0.1606	0.1606	0.1771
		CP	0.9580	0.9420	0.9440	0.9560	0.9560	0.9270	0.9380	0.9130
	β_3	Bias	0.0055	0.1420	0.0235	0.0035	0.0050	0.0072	0.0071	-0.0431
		SD	0.1580	0.2662	0.1634	0.1633	0.1597	0.1580	0.1579	0.1557
		ASE	0.1551	0.2663	0.1575	0.1639	0.1572	0.1555	0.1555	0.1576
		CP	0.9490	0.9270	0.9450	0.9510	0.9450	0.9520	0.9500	0.9470

ρ is the correlation coefficient between X_1 and X_2

of them missing were 16%, 16%, and 4%, respectively. Thus, the size of the CC data set is 320 (64%). In addition, let W denote the hip fracture of mother that is 1 if yes; 0 otherwise. Because W is not significant in logistic regression with the response Y , and W is correlated with X_2 (Spearman’s rank correlation coefficient is 0.13 with p -value 0.025) under the CC data, W is considered as a surrogate variable of X_2 in this study. Moreover, the self-reported risk of fracture was dichotomized to be less than others of the same age ($X_2 = 1$) versus same as or greater than others of the same age ($X_2 = 2$ or 3). The dummy variable DX_2 is used for the dichotomized self-reported risk of fracture, which is 1 if $X_2 = 1$ and 0 otherwise. Age at enrollment (Z) was categorized to be three groups, $Z \leq 60$, $60 < Z \leq 70$, and $Z > 70$. Let DZ_1 and DZ_2 be dummy variables for the categorized age at enrollment. DZ_1 is 1 if $Z \leq 60$ and 0 otherwise, and $DZ_2 = 1$ if $60 < Z \leq 70$ and 0 otherwise. The missingness mechanism of X_1 and X_2 was identified as MAR by Tran et al. (2021). The following logistic regression model is used to fit the data set:

$$P(Y = 1|X_1, X_2, Z, W) = H(\beta_0 + \beta_1 X_1 + \beta_2 DX_2 + \beta_3 DZ_1 + \beta_4 DZ_2). \quad (23)$$

The estimates of β_k s and their corresponding ASEs are given in Table 13. The analysis results indicate that β_k , $k = 0, 1, 2, 3, 4$, are statistically significantly different from zero based on all the estimation methods except β_3 and β_4 based on the

Table 11 Simulation results of scenario 5 ($M = 30$; $\rho = 0.21, 0.51, 0.71$; $n = 1,000$); $\beta = (1, -0.5, 1, \log(2))^T$; $\alpha = (1.15, 0.6, 0.6)^T$; $\gamma = (0.7, -0.2, 0.1, -1.2)^T$; observed selection probabilities: (0.40, 0.23, 0.23, 0.14)

ρ			$\hat{\beta}_F$	$\hat{\beta}_C$	$\hat{\beta}_W$	$\hat{\beta}_V$	$\hat{\beta}_J$	$\hat{\beta}_{M1}$	$\hat{\beta}_{M2}$	$\hat{\beta}_{ME}$
0.21	β_0	Bias	0.0094	0.0821	0.0125	0.0110	0.0072	0.0106	0.0110	-0.1153
		SD	0.1084	0.1709	0.1158	0.1158	0.1118	0.1119	0.1115	0.1008
		ASE	0.1069	0.1698	0.1139	0.1158	0.1090	0.1085	0.1088	0.1089
		CP	0.9450	0.9390	0.9500	0.9550	0.9550	0.9470	0.9510	0.8240
	β_1	Bias	-0.0018	0.0152	-0.0061	-0.0059	0.0069	-0.0024	-0.0041	0.2575
		SD	0.1653	0.2759	0.1965	0.1971	0.1777	0.1828	0.1809	0.1649
		ASE	0.1650	0.2707	0.1954	0.1995	0.1729	0.1720	0.1720	0.1913
		CP	0.9500	0.9520	0.9600	0.9660	0.9440	0.9420	0.9430	0.7570
	β_2	Bias	0.0056	0.0061	0.0121	0.0091	-0.0032	0.0083	0.0078	-0.2392
		SD	0.1216	0.1989	0.1390	0.1385	0.1253	0.1296	0.1281	0.1199
		ASE	0.1210	0.1980	0.1376	0.1413	0.1258	0.1245	0.1251	0.1508
		CP	0.9450	0.9440	0.9460	0.9550	0.9500	0.9370	0.9370	0.6670
β_3	Bias	0.0054	0.1372	0.0098	0.0052	0.0104	0.0058	0.0056	-0.0270	
	SD	0.1601	0.2771	0.1654	0.1659	0.1659	0.1618	0.1616	0.1571	
	ASE	0.1577	0.2722	0.1601	0.1677	0.1602	0.1582	0.1582	0.1572	
	CP	0.9400	0.9430	0.9360	0.9510	0.9340	0.9340	0.9340	0.9470	
0.51	β_0	Bias	0.0078	0.0840	0.0124	0.0116	0.0053	0.0094	0.0026	-0.1489
		SD	0.1067	0.1712	0.1156	0.1163	0.1099	0.1110	0.1088	0.0996
		ASE	0.1077	0.1712	0.1158	0.1181	0.1103	0.1095	0.1092	0.1072
		CP	0.9580	0.9340	0.9560	0.9620	0.9560	0.9510	0.9550	0.7090
	β_1	Bias	-0.0053	0.0103	-0.0160	-0.0150	0.0068	-0.0088	0.0137	0.4810
		SD	0.1898	0.3207	0.2400	0.2391	0.2099	0.2209	0.2131	0.1707
		ASE	0.1898	0.3120	0.2333	0.2404	0.2030	0.2003	0.1999	0.2101
		CP	0.9500	0.9460	0.9510	0.9560	0.9500	0.9210	0.9280	0.3280
	β_2	Bias	0.0123	0.0165	0.0209	0.0182	0.0032	0.0152	0.0024	-0.3039
		SD	0.1327	0.2315	0.1615	0.1599	0.1409	0.1499	0.1439	0.1191
		ASE	0.1344	0.2203	0.1559	0.1613	0.1414	0.1395	0.1387	0.1520
		CP	0.9480	0.9400	0.9370	0.9490	0.9470	0.9230	0.9340	0.4770
β_3	Bias	0.0076	0.1421	0.0169	0.0065	0.0080	0.0088	0.0078	-0.0171	
	SD	0.1632	0.2913	0.1716	0.1693	0.1689	0.1652	0.1646	0.1607	
	ASE	0.1605	0.2775	0.1623	0.1701	0.1627	0.1609	0.1608	0.1603	
	CP	0.9530	0.9190	0.9450	0.9580	0.9430	0.9500	0.9500	0.9590	

Table 11 (continued)

ρ			$\hat{\beta}_F$	$\hat{\beta}_C$	$\hat{\beta}_W$	$\hat{\beta}_V$	$\hat{\beta}_J$	$\hat{\beta}_{M1}$	$\hat{\beta}_{M2}$	$\hat{\beta}_{ME}$
0.71	β_0	Bias	0.0093	0.0831	0.0311	0.0043	0.0031	0.0222	0.0048	- 0.2864
		SD	0.1345	0.2142	0.1718	0.1658	0.1476	0.1556	0.1480	0.1061
		ASE	0.1345	0.2171	0.1607	0.1689	0.1461	0.1406	0.1397	0.1266
		CP	0.9580	0.9370	0.9350	0.9590	0.9540	0.9250	0.9360	0.3620
	β_1	Bias	-0.0029	0.0076	-0.0353	0.0098	0.0166	-0.0247	0.0157	0.6903
		SD	0.2244	0.3728	0.3197	0.3104	0.2599	0.2788	0.2617	0.1681
		ASE	0.2296	0.3779	0.3065	0.3234	0.2605	0.2481	0.2458	0.2268
		CP	0.9570	0.9510	0.9400	0.9520	0.9460	0.9210	0.9380	0.0710
	β_2	Bias	0.0095	0.0170	0.0394	0.0019	-0.0010	0.0284	0.0045	- 0.4092
		SD	0.1590	0.2644	0.2136	0.2054	0.1766	0.1901	0.1784	0.1100
		ASE	0.1600	0.2627	0.1993	0.2116	0.1780	0.1691	0.1676	0.1573
		CP	0.9530	0.9560	0.9310	0.9550	0.9500	0.9200	0.9380	0.1990
β_3	Bias	0.0036	0.1350	0.0187	0.0002	0.0022	0.0062	0.0054	-0.0114	
	SD	0.1581	0.2636	0.1632	0.1607	0.1608	0.1588	0.1587	0.1554	
	ASE	0.1509	0.2587	0.1526	0.1580	0.1528	0.1512	0.1511	0.1511	
	CP	0.9390	0.9250	0.9300	0.9510	0.9390	0.9350	0.9370	0.9360	

ρ is the correlation coefficient between X_1 and X_2

CC method. Based on the SIPW, VL, JCL, MI1, MI2 and MICE methods, the estimates of β_k s and their ASEs were overall quite similar. According to all the estimation methods, the results of testing $\beta_1 = 0$ and $\beta_2 = 0$ and their estimates reveal that women with the history of prior fracture were more likely to have fracture(s) in the first year compared to those without the history of prior fracture, and women with self-reported risk of fracture less than others of the same age were less likely to have fracture(s) in the first year compared to those whose self-reported risk of fracture was the same as or greater than others of the same age. The results of testing $\beta_3 = 0$ and $\beta_4 = 0$ and their estimates based on the SIPW, VL, JCL, MI1, MI2 and MICE estimation methods indicate that women aged 60 years or younger and those older than 60 years and younger than or equal to 70 years were less likely to have fracture(s) in the first year compared to those older than 70 years.

5.2.2 Example 2

The two proposed MI methods are applied to analyze the second real data example, the set of cable television (TV) data collected from the customer survey study of 1,586 residents in three cities in Taiwan (Lee et al. 2011). The satisfaction level of cable TV service is considered as the binary outcome variable, denoted by Y (1 = satisfied; 0 = neutral and dissatisfied). The two covariates with

Table 12 Relative efficiencies in scenario 5 ($M = 30$; $\rho = -0.21, -0.53, -0.71, 0.21, 0.51, 0.71$; $n = 1,000$); $\beta = (1, -0.5, 1, \log(2))^T$; $\alpha = (1.15, 0.6, 0.6)^T$; $\gamma = (0.7, -0.2, 0.1, -1.2)^T$; observed selection probabilities: (0.40, 0.23, 0.23, 0.14)

ρ	β	RE_{C1}	RE_{W1}	RE_{V1}	RE_{J1}	RE_{E1}	RE_{C2}	RE_{W2}	RE_{V2}	RE_{J2}	RE_{E2}	RE_{I2}
-0.21	β_0	2.835	1.048	1.068	1.028	1.238	2.688	0.993	1.013	0.975	1.173	0.948
	β_1	2.152	1.170	1.160	0.901	0.876	2.113	1.149	1.139	0.885	0.860	0.982
	β_2	2.276	1.136	1.127	0.936	2.685	2.299	1.147	1.138	0.945	2.711	1.010
	β_3	3.532	1.106	1.091	1.053	1.026	3.509	1.099	1.083	1.046	1.020	0.994
-0.53	β_0	3.008	1.038	1.047	0.996	0.945	2.943	1.016	1.025	0.975	0.925	0.979
	β_1	2.345	1.155	1.125	0.891	1.734	2.437	1.200	1.169	0.926	1.802	1.039
	β_2	2.294	1.122	1.144	0.910	1.878	2.507	1.226	1.250	0.995	2.052	1.093
	β_3	3.605	1.096	1.092	1.047	1.031	3.604	1.096	1.092	1.047	1.031	1.000
-0.71	β_0	2.689	1.081	1.042	0.974	0.941	2.723	1.094	1.055	0.986	0.953	1.012
	β_1	1.754	1.347	1.325	0.835	1.950	1.900	1.459	1.436	0.904	2.113	1.083
	β_2	2.038	1.275	1.209	0.870	1.497	2.182	1.365	1.295	0.932	1.603	1.071
	β_3	3.639	1.089	1.066	1.020	1.043	3.644	1.091	1.068	1.022	1.045	1.001
0.21	β_0	2.846	1.074	1.071	0.994	1.857	2.864	1.081	1.078	1.000	1.869	1.006
	β_1	2.284	1.156	1.163	0.946	2.797	2.332	1.181	1.188	0.966	2.855	1.021
	β_2	2.348	1.154	1.142	0.931	4.244	2.404	1.182	1.170	0.954	4.346	1.024
	β_3	3.647	1.047	1.051	1.054	0.969	3.657	1.050	1.054	1.057	0.972	1.003
0.51	β_0	2.930	1.089	1.101	0.976	2.586	3.071	1.141	1.153	1.022	2.709	1.048
	β_1	2.107	1.184	1.174	0.902	5.330	2.258	1.269	1.259	0.967	5.713	1.072
	β_2	2.373	1.168	1.141	0.875	4.694	2.600	1.280	1.250	0.959	5.145	1.096
	β_3	3.839	1.086	1.049	1.045	0.954	3.869	1.095	1.057	1.053	0.962	1.008
0.71	β_0	2.137	1.234	1.113	0.882	3.776	2.407	1.390	1.254	0.994	4.255	1.127
	β_1	1.775	1.321	1.231	0.866	6.443	2.023	1.505	1.403	0.987	7.344	1.140
	β_2	1.900	1.277	1.142	0.844	4.860	2.204	1.482	1.325	0.979	5.638	1.160
	β_3	3.473	1.068	1.022	1.024	0.961	3.479	1.070	1.024	1.026	0.963	1.002

ρ is the correlation coefficient between X_1 and X_2

RE_{mk} , $m \in \{C, W, V, J, E\}$, $k = 1, 2$, are the relative efficiencies of the CC, SIPW, VL, JCL, and MICE estimators to MIK estimators for each β_i , $i = 0, 1, 2, 3$

RE_{mk} is the ratio of MSE of m estimator to MSE of MIK estimator for each β_i

$$RE_{mk} = MSE_m / MSE_k, RE_{I2} = MSE_1 / MSE_2, MSE = bias^2 + SD^2$$

missing values are response, denoted by X_1 (1 = Yes; 0 = No), to the question “Have you been given a discount on cable TV?”, and response, denoted by X_2 (0 = the number of children under 12 years old is equal 0; 1 = the number of children under 12 years old is ≥ 1), to the question “How many children under the age of 12 live with you?”. Another non-missing covariate is response, denoted by Z (1 = Yes; 0 = No), to the question “Are you paying for the fourth channel?”. Because the rates of only X_1 missing, only X_2 missing, and both of them missing were 17.9%, 1.3%, and 0.4%, respectively, the CC data set consists of 1,274 respondents (80.3%). A surrogate variable of X_1 and X_2 is response, denoted as

Table 13 Results of logistic regression analysis of GLOW500M data

Variable	Parameters	$\hat{\beta}_C$	$\hat{\beta}_W$	$\hat{\beta}_V$	$\hat{\beta}_J$	$\hat{\beta}_{M1}$	$\hat{\beta}_{M2}$	$\hat{\beta}_{ME}$
Intercept	β_0	-0.7902 (0.2678)	-0.6843 (0.2456)	-0.6863 (0.2477)	-0.7972 (0.2341)	-0.7178 (0.2207)	-0.7248 (0.2226)	-0.7089 (0.2152)
X_1	β_1	0.7265 (0.2895)	0.7179 (0.2991)	0.7337 (0.2977)	0.8330 (0.2671)	0.7717 (0.2545)	0.7973 (0.2673)	0.7304 (0.2681)
DX_2	β_2	-0.7370 (0.3139)	-0.7277 (0.3108)	-0.7448 (0.3105)	-0.6246 (0.2813)	-0.6706 (0.2628)	-0.6918 (0.2685)	-0.5777 (0.2703)
DZ_1	β_3	-0.7180 (0.3936)	-1.1682 (0.3488)	-1.1041 (0.3513)	-1.0699 (0.3413)	-1.1235 (0.3358)	-1.1100 (0.3372)	-1.1743 (0.3292)
DZ_2	β_4	-0.4289 (0.2991)	-0.5871 (0.2475)	-0.5730 (0.2513)	-0.5243 (0.2480)	-0.5823 (0.2430)	-0.5713 (0.2444)	-0.5928 (0.2406)

Values in parentheses are asymptotic standard errors (ASEs)

X_1 is the history of prior fracture (0 = no; 1 = yes)

X_2 is the self-reported risk of fracture (1 = less than others of the same age; 2 = same as others of the same age; 3 = greater than others of the same age)

DX_2 is the dummy variable for the dichotomized self-reported risk of fracture

$DX_2 = 1$ if $X_2 = 1$ and 0 if $X_2 = 2$ or 3.

Z is age at enrollment (integer values from 55 to 90)

DZ_1 and DZ_2 are the dummy variables for the categorized age at enrollment

$DZ_1 = 1$ if $Z \leq 60$ and 0 otherwise

$DZ_2 = 1$ if $60 < Z \leq 70$ and 0 otherwise

W (1 = Yes; 0 = No), to the question “Would you pay extra money for additional channels?”. We fit the following logistic regression model to the data set:

$$P(Y = 1|X_1, X_2, Z, W) = H(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 Z). \tag{24}$$

The multinomial logistic regression model $\log(\pi_j(Y, W, Z)/\pi_4(Y, W, Z)) = \alpha_j + \gamma_{1j}Y + \gamma_{2j}W + \gamma_{3j}Z, j = 1, 2, 3$, is used to examine the effects of Y, W , and Z on the missingness mechanism of X_1 and X_2 , i.e., on their selection probabilities. The p -values of Wald chi-squared tests for the effects of Y (i.e., testing $H_0 : \gamma_{1j} = 0, j = 1, 2, 3$) and Z (i.e., testing $H_0 : \gamma_{3j} = 0, j = 1, 2, 3$) are 0.0021 and < 0.0001 , respectively, so Y and Z are statistically significantly related to the missingness mechanism of X_1 and X_2 and, hence, it is reasonable to assume that X_1 and X_2 are MAR.

Table 14 gives the estimates of $\beta_k, k = 0, 1, 2, 3$, and their corresponding ASEs. The results of testing $\beta_k = 0, k = 0, 1, 2, 3$, are statistically significant except testing $\beta_1 = 0$ by using the MICE estimation method. The estimates of β_1 and β_3 are positive, implying that the respondents were more likely to be satisfied with the cable TV service when they were offered a discount and when they paid for the fourth channel, respectively. Based on all the estimation methods, the estimates of β_2 are

negative, revealing that it is less likely for the respondents to report satisfaction with the cable TV service when there were children under 12 years old in their family. The parameter estimates based on the SIPW, VL, JCL, MI1, MI2, and MICE methods are quite similar and different from those based the CC method. Moreover, the ASEs of the two different types of MI estimators of $\beta_0, \beta_1,$ and β_3 are the smallest in comparison with the other estimation methods, showing the outperformance of the two proposed MI methods and consistence with the simulation results.

Moreover, one can also apply the estimation methods to analyze the cable TV data set with artificial missing values and higher missing rates. To this end, let $(\delta_1^0, \delta_2^0, \delta_3^0, \delta_4^0)$ be indicators for the original missingness statuses of X_1 and X_2 . Let $(\delta_1^*, \delta_2^*, \delta_3^*, \delta_4^*)$ be indicators for artificial missingness statuses of X_1 and X_2 under the assumption of MAR mechanism. Let $(\tilde{\delta}_1, \tilde{\delta}_2, \tilde{\delta}_3, \tilde{\delta}_4)$ be the combination of indicators for the original and artificial missingness statuses of X_1 and X_2 , where $\tilde{\delta}_{ij} = \delta_{i1}^0 \delta_{ij}^* + (1 - \delta_{i1}^0) \delta_{ij}^0, i = 1, 2, \dots, n, j = 1, 2, 3, 4$. The following multinomial logistic regression model

$$\ln \left(\frac{P(\delta_{ik}^* = 1 | Y_i, W_i, Z_i)}{P(\delta_{i4}^* = 1 | Y_i, W_i, Z_i)} \right) = \alpha_k^* + \gamma_1^* Y_i + \gamma_2^* W_i + \gamma_3^* Z_i, k = 1, 2, 3, \quad (25)$$

is applied to generate the data of $\delta_{ij}^*, j = 1, 2, 3, 4$, given (Y_i, W_i, Z_i) , where $\alpha^* = (\alpha_1^*, \alpha_2^*, \alpha_3^*)^T = (1.7, -0.5, 1)^T$ and $\gamma^* = (\gamma_1^*, \gamma_2^*, \gamma_3^*)^T = (-0.7, -0.5, -0.3)^T$. One can obtain the selection probabilities of $(\tilde{\delta}_1, \tilde{\delta}_2, \tilde{\delta}_3, \tilde{\delta}_4)$ as $(0.406, 0.234, 0.231, 0.129)$, which imply that the percentages of complete cases, only X_1 missing, only X_2 missing, and both of them missing are 40.6%, 23.4%, 23.1%, and 12.9%, respectively.

The analysis results of the artificial cable TV data set are given in Table 15. In general, these results have some changes compared to those of the original cable TV data set in Table 14 because of increased missing rates, but they still

Table 14 Results of logistic regression analysis of Cable TV data

Variable	Parameters	$\hat{\beta}_C$	$\hat{\beta}_W$	$\hat{\beta}_V$	$\hat{\beta}_J$	$\hat{\beta}_{M1}$	$\hat{\beta}_{M2}$	$\hat{\beta}_{ME}$
Intercept	β_0	-1.8243 (0.1372)	-1.7830 (0.1306)	-1.7885 (0.1302)	-1.7559 (0.1259)	-1.7657 (0.1257)	-1.7498 (0.1250)	-1.7477 (0.1267)
X_1	β_1	0.3443 (0.1656)	0.3488 (0.1676)	0.3466 (0.1653)	0.3251 (0.1644)	0.3541 (0.1586)	0.3052 (0.1541)	0.2847 (0.1581)
X_2	β_2	-0.4420 (0.1679)	-0.4667 (0.1706)	-0.4482 (0.1679)	-0.5359 (0.1489)	-0.5320 (0.1472)	-0.5381 (0.1481)	-0.5266 (0.1471)
Z	β_3	0.3842 (0.1560)	0.5515 (0.1374)	0.5566 (0.1374)	0.5530 (0.1373)	0.5492 (0.1373)	0.5524 (0.1372)	0.5557 (0.1373)

Values in parentheses are asymptotic standard errors (ASEs)

X_1 (1 = Yes; 0 = No): response to question "Have you been given a discount on cable TV?"

X_2 (0 = the number of children under 12 years old is equal 0; 1 = the number of children under 12 years old is ≥ 1): response to question "How many children under the age of 12 live with you?"

Z (1 = Yes; 0 = No): response to question "Are you paying for the fourth channel?"

represent well the performance properties of the proposed estimation methods. Indeed, the changes are the results of testing $\beta_k = 0, k = 1, 2, 3, 4$. Testing $\beta_1 = 0$ is not statistically significant by using the SIPW, JCL, MI1, MI2, and MICE estimation methods. Testing $\beta_2 = 0$ is not statistically significant by using the CC, SIPW, and VL estimation methods. In addition, the ASEs of all the estimators are increased, especially the CC estimators of all the parameters, and the SIPW and VL estimators of β_1 and β_2 . The ASEs of the two proposed MI estimators are changed less compared to the other estimators. The (absolute) estimates of β_0 and β_3 are increased, and those of β_2 are decreased for almost all the estimation methods. The estimates of β_1 are increased based on the first three estimation methods and decreased based on the last four estimation methods. Again, the parameter estimates based on the MI1 and MI2 estimation methods are changed less compared to the JCL and MICE estimation methods.

In summary, the two real data sets are used to evaluate the applicability of the seven estimation methods and confirm the results in the simulation study section. When the logistic regression model includes all categorical/discrete variables and has two covariates MAR separately or simultaneously, the analytical results are highly consistent with the conclusions in the simulation studies. The biases and ASEs of the MI1, MI2, and JCL estimators are indeed slightly similar and better than those of the SIPW and VL estimators. The results also imply that one may make mistakes in obtaining the fitted logistic regression model if using the CC or MICE (from `mice` package) approach. For instance, in real data example 1, the CC estimation method concludes that DZ_1 and DZ_2 are insignificant; in real data example 2, the MICE estimation method suggests that X_1 is not effective at the significance level 5%. Furthermore, the analysis of the artificial cable TV data example shows that the two proposed methods provide more stable results than the other estimation methods.

Table 15 Results of logistic regression analysis of cable TV data with artificial missing values

Variable	Parameters	$\hat{\beta}_C$	$\hat{\beta}_W$	$\hat{\beta}_V$	$\hat{\beta}_J$	$\hat{\beta}_{M1}$	$\hat{\beta}_{M2}$	$\hat{\beta}_{ME}$
Intercept	β_0	-1.9245 (0.1959)	-1.8308 (0.1562)	-1.8504 (0.1551)	-1.7588 (0.1343)	-1.8015 (0.1322)	-1.7884 (0.1303)	-1.7943 (0.1343)
X_1	β_1	0.4529 (0.2256)	0.4196 (0.2304)	0.4568 (0.2251)	0.2101 (0.1925)	0.3152 (0.1731)	0.3078 (0.1666)	0.2420 (0.1762)
X_2	β_2	-0.4187 (0.2409)	-0.4486 (0.2488)	-0.4289 (0.2417)	-0.4544 (0.1805)	-0.4154 (0.1668)	-0.4450 (0.1703)	-0.3707 (0.1763)
Z	β_3	0.5046 (0.2187)	0.5485 (0.1390)	0.5604 (0.1395)	0.5597 (0.1372)	0.5561 (0.1372)	0.5547 (0.1374)	0.5655 (0.1371)

Values in parentheses are asymptotic standard errors (ASEs)

X_1 (1 = Yes; 0 = No): response to question “Have you been given a discount on cable TV?”.

X_2 (0 = the number of children under 12 years old is equal 0; 1 = the number of children under 12 years old is ≥ 1): response to question “How many children under the age of 12 live with you?”.

Z (1 = Yes; 0 = No): response to question “Are you paying for the fourth channel?”

6 Conclusion

Two different types of MI methods have been proposed to estimate the parameters of logistic regression with covariates missing separately or simultaneously. Based on the idea of Wang and Chen (2009), for each type MI estimation method we have proposed three empirical CDFs to generate the random values for missing data and estimated the logistic regression parameters by using the estimating equations of Fay (1996), which are more convenient in practice than those of Rubin (1987) because they are solved only once. The simulation studies have showed that the two proposed MI methods' performances were comparable with the JCL method, but they had shorter computing time than the JCL and MICE estimation methods, and are easily implemented. The two proposed MI methods overall outperformed the CC, SIPW, VL, and MICE methods. Two real data sets have been used to illustrate the practical use of the two proposed MI methods.

Although we have focused on the case where covariates are categorical/discrete, one can also consider the case of continuous covariates by using the non-parametric kernel approach of Wang and Chen (2009) to construct the empirical CDFs. For example, assume that X_1 and X_2 are continuous covariate vectors and MAR separately or simultaneously, Z is a categorical/discrete covariate vector and always observed, W_1 and W_2 are categorical/discrete surrogate vectors of X_1 and X_2 , respectively, and, hence, $V = (Z^T, W^T)^T$ is a categorical/discrete covariate vector. The MI2 method still works for the case. For the MI1 method, to impute the missing values of X_{1i} and X_{2i} , one can construct the empirical CDFs of X_{1i} given (X_{2i}, Y_i, V_i) and X_{2i} given (X_{1i}, Y_i, V_i) as follows:

$$\begin{aligned} \tilde{F}_{X_{1i}}^*(x_1|X_{2i}, Y_i, V_i) &= \sum_{k=1}^n \left(\frac{\delta_{k1}I(Y_k = Y_i, V_k = V_i)\mathcal{K}_{\mathcal{H}_1}(X_{2k} - X_{2i})}{\sum_{s=1}^n \delta_{s1}I(Y_s = Y_i, V_s = V_i)\mathcal{K}_{\mathcal{H}_1}(X_{2s} - X_{2i})} \right) I(X_{1k} \leq x_1), \\ \tilde{F}_{X_{2i}}^*(x_2|X_{1i}, Y_i, V_i) &= \sum_{k=1}^n \left(\frac{\delta_{k1}I(Y_k = Y_i, V_k = V_i)\mathcal{K}_{\mathcal{H}_2}(X_{1k} - X_{1i})}{\sum_{s=1}^n \delta_{s1}I(Y_s = Y_i, V_s = V_i)\mathcal{K}_{\mathcal{H}_2}(X_{1s} - X_{1i})} \right) I(X_{2k} \leq x_2). \end{aligned}$$

Here \mathcal{H}_s is a $p_s \times p_s$, $s = 1, 2$, symmetric positive definite matrix and depends on n for $p_1 = r_2 = \text{length}(X_2)$ and $p_2 = r_1 = \text{length}(X_1)$. $\mathcal{K}_{\mathcal{H}_s}(\mathbf{u}) = |\mathcal{H}_s|^{1/2}\mathcal{K}_s(\mathcal{H}_s^{-1/2}\mathbf{u})$ for $\mathcal{K}_s(\cdot)$ being a p_s -variate kernel with $\int \mathcal{K}_s(\mathbf{u})d\mathbf{u} = 1$ and $\mathcal{H}_s^{1/2}$ the bandwidth matrix. $\tilde{F}_{X_i}^*(\mathbf{x}|Y_i, V_i)$ is equal to $\tilde{F}_{X_i}(\mathbf{x}|Y_i, V_i)$ in (17). We can use the MI1 procedure in Secti. 4.1 to estimate the parameters. The simulation studies demonstrate the two proposed MI methods still work well for this case; simulation results are not reported.

Furthermore, to build the MI1 method for the case where both X_1 and X_2 are continuous covariate vectors and MAR separately or simultaneously, Z is a continuous covariate vector, and W_1 and W_2 are continuous surrogate vectors of X_1 and X_2 , respectively, one can construct the following empirical CDFs of X_{1i} given (X_{2i}, Y_i, V_i) , X_{2i} given (X_{1i}, Y_i, V_i) , and X_i given (Y_i, V_i) :

$$\begin{aligned} \tilde{F}_{X_{2i}}^{**}(x_1 | X_{2i}, Y_i, V_i) &= \sum_{k=1}^n \left(\frac{\delta_{k1} I(Y_k = Y_i) \mathcal{K}_{\mathcal{H}_1}(X_{2k} - X_{2i}, \mathbf{V}_k - \mathbf{V}_i)}{\sum_{s=1}^n \delta_{s1} I(Y_s = Y_i) \mathcal{K}_{\mathcal{H}_1}(X_{2s} - X_{2i}, \mathbf{V}_s - \mathbf{V}_i)} \right) I(X_{1k} \leq x_1), \\ \tilde{F}_{X_{2i}}^{**}(x_2 | X_{1i}, Y_i, V_i) &= \sum_{k=1}^n \left(\frac{\delta_{k1} I(Y_k = Y_i) \mathcal{K}_{\mathcal{H}_2}(X_{1k} - X_{1i}, \mathbf{V}_k - \mathbf{V}_i)}{\sum_{s=1}^n \delta_{s1} I(Y_s = Y_i) \mathcal{K}_{\mathcal{H}_2}(X_{1s} - X_{1i}, \mathbf{V}_s - \mathbf{V}_i)} \right) I(X_{2k} \leq x_2), \\ \tilde{F}_{X_i}^{**}(\mathbf{x} | Y_i, V_i) &= \sum_{k=1}^n \left(\frac{\delta_{k1} I(Y_k = Y_i) \mathcal{K}_{\mathcal{H}_3}(\mathbf{V}_k - \mathbf{V}_i)}{\sum_{s=1}^n \delta_{s1} I(Y_s = Y_i) \mathcal{K}_{\mathcal{H}_3}(\mathbf{V}_s - \mathbf{V}_i)} \right) I(\mathbf{X}_k \leq \mathbf{x}). \end{aligned}$$

Here \mathcal{H}_s is a $q_s \times q_s$, $s = 1, 2, 3$, symmetric positive definite matrix and depends n for $q_1 = r_2 + \text{length}(\mathbf{V})$, $q_2 = r_1 + \text{length}(\mathbf{V})$, and $q_3 = \text{length}(\mathbf{V})$. $\mathcal{K}_{\mathcal{H}_s}(\mathbf{u}) = |\mathcal{H}_s|^{1/2} \mathcal{K}_s(\mathcal{H}_s^{-1/2} \mathbf{u})$ for $\mathcal{K}_s(\cdot)$ being a q_s -variate kernel with $\int \mathcal{K}_s(\mathbf{u}) d\mathbf{u} = 1$ and $\mathcal{H}_s^{1/2}$ the bandwidth matrix. Similarly, to develop the MI2 method for this case, one can form the empirical CDFs of X_{1i} given (X_{2i}, Y_i, V_i) , X_{2i} given (X_{1i}, Y_i, V_i) , and X_i given (Y_i, V_i) as follows:

$$\begin{aligned} \tilde{F}_{X_{1i}}^{**}(x_1 | Y_i, V_i) &= \sum_{k=1}^n \left(\frac{(\delta_{k1} + \delta_{k3}) I(Y_k = Y_i) \mathcal{K}_{\mathcal{H}}(\mathbf{V}_k - \mathbf{V}_i)}{\sum_{s=1}^n (\delta_{s1} + \delta_{s3}) I(Y_s = Y_i) \mathcal{K}_{\mathcal{H}}(\mathbf{V}_s - \mathbf{V}_i)} \right) I(X_{1k} \leq x_1), \\ \tilde{F}_{X_{2i}}^{**}(x_2 | Y_i, V_i) &= \sum_{k=1}^n \left(\frac{(\delta_{k1} + \delta_{k2}) I(Y_k = Y_i) \mathcal{K}_{\mathcal{H}}(\mathbf{V}_k - \mathbf{V}_i)}{\sum_{s=1}^n (\delta_{s1} + \delta_{s2}) I(Y_s = Y_i) \mathcal{K}_{\mathcal{H}}(\mathbf{V}_s - \mathbf{V}_i)} \right) I(X_{2k} \leq x_2), \\ \tilde{F}_{X_i}^{**}(\mathbf{x} | Y_i, V_i) &= \sum_{k=1}^n \left(\frac{\delta_{k1} I(Y_k = Y_i) \mathcal{K}_{\mathcal{H}}(\mathbf{V}_k - \mathbf{V}_i)}{\sum_{s=1}^n \delta_{s1} I(Y_s = Y_i) \mathcal{K}_{\mathcal{H}}(\mathbf{V}_s - \mathbf{V}_i)} \right) I(\mathbf{X}_k \leq \mathbf{x}). \end{aligned}$$

Here \mathcal{H} is a $q_3 \times q_3$ symmetric positive definite matrix and depends n . $\mathcal{K}_{\mathcal{H}}(\mathbf{u}) = |\mathcal{H}|^{1/2} \mathcal{K}(\mathcal{H}^{-1/2} \mathbf{u})$ for $\mathcal{K}(\cdot)$ being a q_3 -variate kernel with $\int \mathcal{K}(\mathbf{u}) d\mathbf{u} = 1$ and $\mathcal{H}^{1/2}$ the bandwidth matrix.

Finally, as an extension, the two proposed MI methods can be applied to estimate the parameters of logistic regression with outcome and covariates missing separately or simultaneously, and to estimate the parameters of a multinomial logit model with the same missing issue. These tasks are future research.

Acknowledgements The authors thank two referees and an Associate Editor for their constructive comments that improved the presentation. The research of S.M. Lee and T.N. Le was supported by Ministry of Science and Technology (MOST) Grant of Taiwan, ROC, MOST-109-2118-M-035-002-MY3.

References

Breslow NE, Cain KC (1988) Logistic regression for two-stage case-control data. *Biometrika* 75:11–20
 Buuren SV, Groothuis-Oudshoorn K (2011) Mice: multivariate imputation by chained equations in R. *J Stat Softw* 45(3):1–67
 Dong Y, Peng CYJ (2013) *Principled missing data methods for researchers*. Springer, Berlin
 Fay RE (1996) Alternative paradigms for the analysis of imputed survey data. *J Am Stat Assoc* 91:490–498

- Horvitz DG, Thompson DJ (1952) A generalization of sampling without replacement from a finite universe. *J Am Stat Assoc* 47:663–685
- Hosmer DW, Lemeshow S, Sturdivant RX (2013) *Applied logistic regression*, 3rd edn. Wiley, New York
- Hsieh SH, Lee SM, Shen PS (2010) Logistic regression analysis of randomized response data with missing covariates. *J Stat Plann Infer* 140:927–940
- Hsieh SH, Li CS, Lee SM (2013) Logistic regression with outcome and covariates missing separately or simultaneously. *Comput Stat Data Anal* 66:32–54
- Jiang W, Josse J, Lavielle M, Group T (2020) Logistic regression with missing covariates|parameter estimation, model selection and prediction within a joint-modeling framework. *Comput Stat Data Anal* 145:106907
- Lee SM, Gee MJ, Hsieh SH (2011) Semiparametric methods in the proportional odds model for ordinal response data with missing covariates. *Biometrics* 67:788–798
- Lee SM, Hwang WH, de Dieu Tapsoba J (2016) Estimation in closed capture-recapture models when covariates are missing at random. *Biometrics* 72:1294–1304
- Lee SM, Li CS, Hsieh SH, Huang LH (2012) Semiparametric estimation of logistic regression model with missing covariates and outcome. *Metrika* 75:621–653
- Lee SM, Lukusa TM, Li CS (2020) Estimation of a zero-inflated Poisson regression model with missing covariates via nonparametric multiple imputation methods. *Computat Stat* 35:725–754
- Lipsitz SR, Parzen M, Ewell M (1998) Inference using conditional logistic regression with missing covariates. *Biometrics* 54:295–303
- Little RJ (1992) Regression with missing X's: a review. *J Am Stat Assoc* 87:1227–1237
- Little RJ, Rubin DB (2019) *Statistical analysis with missing data*, 3rd edn. Wiley, New York
- Lukusa TM, Lee SM, Li CS (2016) Semiparametric estimation of a zero-inflated Poisson regression model with missing covariates. *Metrika* 79:457–483
- Pahel BT, Preisser JS, Stearns SC, Rozier RG (2011) Multiple imputation of dental caries data using a zero-inflated Poisson regression model. *J Public Health Dent* 71:71–78
- Rubin DB (1976) Inference and missing data. *Biometrika* 63:581–592
- Rubin DB (1987) *Statistical analysis with missing data*. Wiley, New York
- Rubin DB (1996) Multiple imputation after 18+ years. *J Am Stat Assoc* 91:473–489
- Rubin DB, Schenker N (1986) Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *J Am Stat Assoc* 81:366–374
- Tran PL, Le TN, Lee SM, Li CS (2021) Estimation of parameters of logistic regression with covariates missing separately or simultaneously. *Communications in statistics - Theory and methods*, in press
- Wang CY, Chen JC, Lee SM, Ou ST (2002) Joint conditional likelihood estimator in logistic regression with missing covariate data. *Statistica Sinica* 12:555–574
- Wang CY, Wang S, Zhao LP, Ou ST (1997) Weighted semiparametric estimation in regression analysis with missing covariate data. *J Am Stat Assoc* 92:512–525
- Wang D, Chen SX (2009) Empirical likelihood for estimating equations with missing values. *Ann Stat* 37:490–517
- Wang S, Wang CY (2001) A note on kernel assisted estimators in missing covariate regression. *Statistics and Probability Letters* 55:439–449
- White IR, Royston P, Wood AM (2011) Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 30:377–399
- Zhao LP, Lipsitz S (1992) Designs and analysis of two-stage studies. *Stat Med* 11:769–782

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Shen-Ming Lee¹  · Truong-Nhat Le^{1,2}  · Phuoc-Loc Tran³  · Chin-Shang Li⁴ 