

Research Article

Evaluation of Four Multiple Imputation Methods for Handling Missing Binary Outcome Data in the Presence of an Interaction between a Dummy and a Continuous Variable

Sara Javadi ¹, Abbas Bahrapour ², Mohammad Mehdi Saber ³, Behshid Garrusi ⁴,
and Mohammad Reza Baneshi ²

¹Department of Biostatistics and Epidemiology, School of Public Health, Kerman University of Medical Sciences, Kerman, Iran

²Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

³Department of Statistics, Higher Education Center of Eghlid, Eghlid, Iran

⁴Kerman Neuroscience Research Center, Institute of Neuropharmacology, Kerman University of Medical Sciences, Kerman, Iran

Correspondence should be addressed to Mohammad Reza Baneshi; rbaneshi2@gmail.com

Received 15 October 2020; Revised 9 March 2021; Accepted 22 April 2021; Published 18 May 2021

Academic Editor: Marek T. Malinowski

Copyright © 2021 Sara Javadi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Multiple imputation by chained equations (MICE) is the most common method for imputing missing data. In the MICE algorithm, imputation can be performed using a variety of parametric and nonparametric methods. The default setting in the implementation of MICE is for imputation models to include variables as linear terms only with no interactions, but omission of interaction terms may lead to biased results. It is investigated, using simulated and real datasets, whether recursive partitioning creates appropriate variability between imputations and unbiased parameter estimates with appropriate confidence intervals. We compared four multiple imputation (MI) methods on a real and a simulated dataset. MI methods included using predictive mean matching with an interaction term in the imputation model in MICE (MICE-interaction), classification and regression tree (CART) for specifying the imputation model in MICE (MICE-CART), the implementation of random forest (RF) in MICE (MICE-RF), and MICE-Stratified method. We first selected secondary data and devised an experimental design that consisted of 40 scenarios ($2 \times 5 \times 4$), which differed by the rate of simulated missing data (10%, 20%, 30%, 40%, and 50%), the missing mechanism (MAR and MCAR), and imputation method (MICE-Interaction, MICE-CART, MICE-RF, and MICE-Stratified). First, we randomly drew 700 observations with replacement 300 times, and then the missing data were created. The evaluation was based on raw bias (RB) as well as five other measurements that were averaged over the repetitions. Next, in a simulation study, we generated data 1000 times with a sample size of 700. Then, we created missing data for each dataset once. For all scenarios, the same criteria were used as for real data to evaluate the performance of methods in the simulation study. It is concluded that, when there is an interaction effect between a dummy and a continuous predictor, substantial gains are possible by using recursive partitioning for imputation compared to parametric methods, and also, the MICE-Interaction method is always more efficient and convenient to preserve interaction effects than the other methods.

1. Introduction

The need to adequately deal with missing data is a practical challenge for researchers when analyzing epidemiological data. A well-known approach (and the default in most statistical packages) to deal with the missing data problems is complete case analysis (CCA), which omits subjects with

missing values from the analysis. In some cases, such analyses are inefficient, since they sacrifice information from partially observed responses, and even in the worst situations, they may result in biased inferences about the parameters of interest [1]. An alternative to CCA is MI, which creates copies of the dataset, replacing the missing values in each dataset with independent random draws from the

predictive distribution of the missing values under a specific model (the imputation model). Each dataset is analyzed separately and the results of M datasets are combined with a set of simple rules. The corresponding estimates (M point and M variance estimates) are combined according to Rubin's combination rule [2].

Using Rubin's rule [2], the reasons for missing data are classified as missing completely at random (MCAR) when the probability of missingness is independent of the observed and unobserved data, missing at random (MAR) if the probability of missingness is independent of the unobserved data after conditioning on observed data, and missing not at random (MNAR), where the probability of missingness is dependent on unobserved data even after conditioning on observed data [3–6].

MICE, also called fully conditional specification (FCS), is commonly used for imputing missing data. The MICE method specifies the univariate distribution of each incomplete variable conditional on all other variables and creates imputations per variable. The MICE algorithm is a Gibbs sampler, a Bayesian simulation approach that generates random draws from the posterior distribution and conducts univariate imputations sequentially until convergence [7–10].

In the MICE algorithm, the set of conditional distributions may not correspond to any joint distribution since the users specify univariate conditional distributions, and so the joint distribution may not actually exist. Despite this theoretical drawback of the MICE method, the simulation study suggests this imputation method performs well in practice [11, 12]. There are a number of software packages available to impute missing data using MICE methods. These include "IVEware" in SAS [11], "mice" [12, 13] and "mi" in R [14], and "mi" and "ice" in STATA [15].

In the MICE strategy, when the data include an interaction effect, the interaction can be modeled by appropriate models manually and by imputing the missing values in separate subgroups of the data. The default setting in implementation of MICE is for imputation models to include variables as linear terms only with no interactions, but omission of important nonlinear terms may lead to biased results [16].

Motivated by these challenges, several authors have developed more flexible techniques that can handle missing values in the presence of interactions easily. Automatic Interaction Detection is one of the first implementations of recursive partitioning [17]. Besides the fact that the recursive partitioning technique finds the split that is most predictive of the response variable by searching through all predictor variables, they model the interaction structure in the dataset by splitting a dataset into increasingly homogeneous subgroups sequentially. In other words, since splits are conditional on previous splits, possible interactions are automatically detected. Others have used an approach combining of recursive partitioning with imputation methods [3, 18]. They used CART as an imputation engine in MICE. CART in MICE is available as an option in "mice" in R.

CART methods have properties that make them attractive for imputation: they are robust against outliers, can deal with multicollinearity and skewed distributions, and are flexible enough to fit interactions and nonlinear relations. Indeed, many aspects of model fitting have been automated, so there is little tuning needed by the imputer [18].

However, to our knowledge, few evaluations have been done in simulation contexts; even fewer evaluations have been done on data, in the presence of interaction with mixtures of binary and continuous predictors. This lack of comparisons makes it difficult, if not impossible, to assess the relative merits of each procedure.

Thus, to find out when the data include a mixture of continuous and binary predictors in the presence of an interaction, which method is best to impute the missing values in the binary response under MAR and MCAR missing mechanisms, we need to evaluate the performance of the four methods for handling missing data in a real and simulated dataset. This evaluation can be further justified through the following reasons: first, most of the studies that evaluated and compared the performance of these four methods have not included the mixture of continuous and binary predictors with binary outcome. Second, no study has yet been made due to the challenge that MICE-Interaction and MICE-Stratified will result in proper imputation in datasets with interaction between a dummy and continuous predictor. Third, it is not clear what proportion of missing data via MAR and MCAR correspond to a certain amount of bias in the results, so even if we accept that one method has a kind of superiority over others, it is desirable to know how much of the bias can be compensated by that method.

Regarding the above explanations, this study was carried out to evaluate the performance of four MI methods in real and simulated data. At first, we performed MICE-Stratified, filling in values separately for the two subgroups defined by the value of the binary variable in each dataset. Then, we used three MI methods. First, we used MICE-Interaction method. For MICE-interaction, we included the interaction term as just another variable in the predictor matrix and used the default predictor matrix in the *mice* function. Then, we used CART for specifying the imputation model in MICE (MICE-CART), and the implementation of random forest (RF) in MICE (MICE-RF). We decline to provide formulae or mathematical details in this article because we want to focus more on simulation.

The paper is organized as follows. In Section 2, MICE-Interaction and MICE-Stratified are first elaborated. In continuation of this section, two main recursive partitioning techniques are considered, namely, CART and RF. Subsequently, incorporation of the CART method and RF method in the MICE framework is presented. In Section 3, we implement a secondary empirical study to investigate the performance of the discussed methods. In Section 4, a simulation study is carried out to investigate which of the discussed methods are convenient to preserve the interaction effects. The results from both studies are discussed in Section 5; at the end of the study, some final conclusions are given.

2. Methods

2.1. MICE Approach. Suppose that data are presented as $X = (X^{\text{miss}}, X^{\text{obs}})$, where X^{miss} is included of the columns with at least one missing value, X^{obs} is included the columns of X that are completely observed, X_j is the j th column of X , X_j^{miss} is the missing values in the j th column of the X^{miss} , Z is the currently imputed data matrix X , and k is the number of partially observed variables. Suppose that X^{miss} is ordered in non-decreasing numbers of missing values in each column. Define X_{-j}^{miss} equal to the matrix X with its j th column removed. Thus, we will have the following algorithm to show the implementation of standard MICE:

- (1) To fill in the initial values for the missing values, define a matrix Z equal to X^{obs} ; for each X_j^{miss} , all the X_j^{miss} values are initially filled in by random draws from the predictive distribution conditional on Z , and attach the imputed version of X_j^{miss} to Z prior to incrementing j .
- (2) For $j = 1, \dots, k$, replace the missing values of X_j with random draws from the predictive distribution conditional on X_{-j}^{miss} .
- (3) Repeat steps 1 and 2 for a number of iterations. This procedure is performed for each variable with at least one missing value, yielding one complete dataset.
- (4) Repeat steps 1–3 a number of times (M), resulting in M imputed datasets that are available for analysis. Each of the M imputed datasets is analyzed separately. In the next step, the results are combined using Rubin's rules. It is standard to use generalized linear models as the basis of the posterior predictive distribution draws in steps 1 and 2.

For MICE-Interaction method, the interaction is included in the predictor matrix. In the other words, the interaction term is included as just another variable in the imputation model and then the *mice* function is used since this method performs better than or equivalent to all other methods considered (the “mice” package in R does this) [3, 7, 8, 19].

For MICE-Stratified method, the datasets are stratified by dummy variable and then MICE method is used to impute missing values, in the presence of interaction effect between a dummy and a continuous variable.

An alternative approach is described in [20]. They defined a new class of nonparametric multiple imputation methods based on the CART or RF algorithm. These two methods fall into the umbrella concept of “recursive partitioning,” which allows for the modelling of internal interactions in the data by sequentially partitioning the dataset into homogeneous subsets. Some researchers used the tree package and showed that the CART results for recovering interactions were uniformly better than standard techniques [18]. Shah and coworkers applied random forest techniques to both continuous and categorical outcomes, which produced more efficient estimates than standard procedures [21]. A similar set of routines building on the *rpart* [22] and *randomForest* [23] packages were developed by Doove and

coworkers [22]. Methods CART and RF are part of mice package.

2.2. Imputation by MICE-CART. CART is a nonparametric recursive partitioning imputation method that provides the results as a tree structure. The root node is at the top of the tree, which includes all members. It follows by exploring the data to find the best variable and a cut-off that best separates the subjects into two child nodes. Subgroups are made by the optimal split according to a measure of homogeneities such as the Gini index [24]. Partitioning of each child node continues until some certain stopping criterion has been reached, e.g., a predetermined number of observations in the final subsets [25]. Therefore, tree models easily reveal the interaction between independent variables [26].

Suppose that the data are presented as a matrix $X = (X^{\text{obs}}, X^{\text{miss}})$, where X^{obs} contains the columns of X with observed data for all variables and X^{miss} is the part of X that has at least one missing datum. Similar to MICE, the variable with missing data is considered as a dependent variable. The following steps indicate an implementation of CART in MICE:

- (1) For each variable with missing data ($j = 1, \dots, k$), fill in the initial values \hat{X}_j by random draws from X_j^{obs} , and update the X^{obs} matrix (shown by Z).
- (2) Fit the CART using each X_j^{miss} as outcome, and Z as predictor variables; only subjects with observed values on X_j^{miss} are used in this process X^{obs} .
- (3) For subjects in X_j^{miss} , find the terminal node; they end up according to the fitted tree in step 2; and one observed value on X_j^{miss} is randomly selected from the subset in this node and used for imputation.
- (4) Repeat steps 2 and 3 for a number of iterations. This procedure is performed for each variable with at least one missing value, yielding one complete dataset.
- (5) Repeat steps 1–4 a number of times (M), resulting in M imputed datasets.

To construct the tree, a minimum leaf size of 5 considered, with the deviance of less than 0.0001. Additionally, the response variable (Y) was included as a predictor in the imputation models to impute each incomplete variable [3, 18, 20].

2.3. Imputation by MICE-RF. The algorithm needed for RF imputation is a modification of the discussed CART algorithm. The first two steps are replaced by a construction of k bootstrapped datasets, k being the number of trees in the forest, and the fitting of k tree models. Optionally, each tree can be fitted using the full bootstrapped dataset or randomly selecting the input variables. To avoid reduced variability by imputing based on an averaged tree, possibly due to the higher stability of the individual trees, the imputed value is randomly selected from the union of the k donor pools. For more details on the algorithm, see [22].

3. Empirical Study

We used empirical data from a population-based study collected in southeastern Iran [27]. The data was initially collected to address the main variables that encourage participants to change their body using a variety of methods such as diet, exercise, or drugs. A multistage sampling was adopted where the area was stratified into 10 blocks. From each block, approximately 120 households were selected. In each block, the first household was selected at random. The remaining blocks were selected following a systematic sampling approach. In each household, only one person, aged 14–55, was interviewed.

3.1. Dependent and Independent Variables. The outcome variable of interest was whether participants had tried to change their body using a variety of methods such as diet, surgery, or drugs. Independent variables include BMI (body mass index), body-esteem scale (BE), perceived sociocultural pressure scale (PSPS), physical appearance comparison scale (PACS), and Gender. Our data did not involve any missing values. These variables were measured with standard and validated questionnaires. Logistic regression analysis showed that BMI, BE, PACS4, PSPS, and Gender were independent variables that could influence eating disorders. More details are provided in [27].

In this study, we proposed the following complete-data model (1), which contains all five main variables and an interaction term between a dummy and a continuous variable to guide the analysis of the datasets.

$$\text{Logit}[P(Y = 1)] = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_1 * X_5, \quad (1)$$

where Y is the binary outcome variable, X_1 is BMI, X_2 is PSPS, X_3 is BE, X_4 is PACS4, X_5 is Gender, and $X_1 * X_5$ is the interaction between BMI and Gender.

3.2. Simulation Study for Real Data (Generation of Missing Data). In this section, the simulation was restricted to generating missing data in variable Y . In fact, for the experimental data, the following steps were repeated 1000 times:

Step 1. First, We randomly drew 700 observations 300 times with replacement from each of the datasets. 10% to 50% univariate missing data were created in the outcome variable via MCAR and MAR mechanisms. MAR data was generated in variable Y in a way that probability of becoming missing depended on X_3 and X_4 .

Step 2. Using MI method: the four considered MI methods were MICE-Interaction, MICE-CART, MICE-RF, and MICE-Stratified. The independent variables used in all the methods were Gender, BMI, BE, PACS4, PSPS, and the interaction between Gender and BMI. Finally, the number of imputations was set to 5 [3, 28].

Step 3. Recalculation of the coefficients and other indexes [5] from complete datasets, which is intended to represent how MI was able to reduce the effect of missing data on the estimations.

Step 4. There are several measures that may inform us about the statistical validity of a particular procedure. This step includes calculations of raw bias (RB), percent bias (PB), coverage rate (CR), model-based and empirical SE, and estimated proportion of the variance attributable to the missing data ($\hat{\lambda}$). The raw bias, which is the average difference between the true value of the parameter being estimated from the real data, and the value of the estimation after imputation, should be close to 0. Bias can also be expressed as percent bias (PB). For acceptable performance, we use an upper limit for PB of 5%. In more detail, the grades of PB are defined as negligible (0%–5%), minimal (5%–10%), moderate (10%–20%), heavy (20%–30%), and severe (>30%) [29, 30]. The coverage is the percentage of cases where the estimand, the true value of the parameter being estimated [20], is located within the 95% confidence interval around the estimate, which should be 95% or higher. The width of the 95% confidence interval should be as small as possible (as long as the coverage does not fall below 95%) and is an indicator of statistical efficiency. However, it is important to evaluate the coverage with other measures, because high variances can lead to higher coverage rates. We regard the performance of the interval procedure to be poor if its coverage drops below 90% [31]. Model-based SE is the mean of the SE estimated across simulations and empirical SE is the SD of the estimates across simulations. The two should be similar and should match if the Rubin's variance estimate with a specific MI model is unbiased. Lastly, $\hat{\lambda}$ is the estimated proportion of the variation attributable to the missing data, and $\hat{\lambda}$ is an indicator of the severity of the missing data problem (the computation of RB, PB, and λ is based on equations (2)–(4)). The mice package of R was used to calculate these measurements.

$$\text{RB} = E(\bar{Q} - Q), \quad (2)$$

$$\text{PB} = \left| 100 \left[\frac{E(\bar{Q} - Q)}{Q} \right] \right|, \quad (3)$$

$$\lambda = \frac{B + (B/M)}{T}. \quad (4)$$

A scientific estimand Q is a quantity of scientific interest that we can calculate if we would observe the entire population. T is the total variance of \bar{Q} , and hence of $(Q - \bar{Q})$ if \bar{Q} is unbiased. B is the extra variance caused by the fact that there are missing values in the sample, and B/M is the extra simulation variance caused by the fact that the imputation reputation M is finite [3].

4. Simulation Study

This simulation included the generation of artificial datasets and then creating missing data in the binary outcome variable. The following steps are performed to examine the imputation methods.

Step 1. In this simulation study, we generated a finite population of size $N = 700$ from the binary variable outcome Y . The artificial data include one binary predictor variable and four continuous variables that the binary variable was uncorrelated from the other variables. The number of continuous variables is kept higher than the number of categorical variables due to the fact that the simulation is aimed to be similar to our real data. One binary variable was randomly drawn from a binomial distribution. The marginal distribution of X_5 is Bernoulli (0.6). Moreover, four continuous predictors (X_1, X_2, X_3 and X_4) were randomly drawn from a multivariate normal distribution, where the first two predictor variables (X_1 and X_2) had a pairwise correlation of $r = 0.5$ and the last two predictor variables (X_3 and X_4) had pairwise correlations of $r = 0.3$. The binary response is modeled using GLM depending on various binary and continuous predictors. The dichotomization of X_5 is based on the following criteria:

$$P(X_5 = 1) = P(X_5 \leq \mu_5) = 0.6, \quad (5)$$

where 0.6 is the mean value of X_5 . By defining and standardizing $\text{logit}[P(Y_i = 1)] = \beta_1 X_{i1} + \dots + \beta_{p-1} X_{i,p-1} + \beta_p X_{i1} X_{i,p-1}$, $\theta = \beta_{\text{true}} = (0.25, 0.5, 0.5, 0.5, 0.5, 1)$, $p = 6$, $I = 1, \dots, N$.

We have used the simulation design that combined sampling and missing data mechanisms. To achieve this simulation design, we first generated 1000 sets of data based on the described characteristics and then created missing values one time for each complete dataset. Steps 2 and 3 are the same as those described in the real data simulation section. Various scenarios have been considered, including two missing mechanisms (MAR and MCAR), five missing proportions (10%, 20%, 30%, 40%, and 50%), and four different imputation methods (MICE-Interaction, MICE-CART, MICE-RF, and MICE-Stratified), which lead to a combination of $2 \times 5 \times 4 = 40$ scenarios. According to the above, a total of 1000 repetitions were performed. To compare different scenarios, the average measures for the scenarios were calculated.

5. Results

5.1. Results for Real Data. For the real dataset, we present only the results with respect to the interaction effects in the form of plots. The results of the imputation methods under the missing mechanism of MCAR and MAR are shown in Figures 1–6. They are separated by different measurements.

5.1.1. Raw Bias (RB). All the imputation methods used to estimate the interaction were almost biased towards the null except the MICE-Stratified method. However, the bias values of MICE-Stratified method were lower than 0.1 at all the missing percentages. The MICE-Interaction and MICE-CART methods were less biased than the other two methods. We found that the bias values become higher with increasing missing percentages (Figure 1).

5.1.2. Percent Bias (PB). The MICE-Interaction and MICE-CART methods were less biased than the other two methods at almost all scenarios. The MICE-Stratified method produced the highest PB at all missing percentages (Figure 2).

5.1.3. Coverage Rate (CR). Among imputation methods, the MICE-RF method had the highest coverage rates at all missing rates and both mechanisms and the coverage rate of the MICE-Stratified method was very low (Figure 3).

5.1.4. Model and Empirical SE. At all the MI approaches, the empirical and model-based standard errors are almost identical for the interaction effect (Figures 4 and 5).

5.1.5. The Proportion of Variation Attributable to the Missing Data. In general, λ values for all methods were less than 0.5. The MICE-Stratified and MICE-CART methods had lower λ value for interaction than the others (for instance, the MICE-Stratified led to the lowest bias in the MCAR missing mechanism, at all the missing percentages), which showed that the MICE-RF and MICE-Interaction methods indicated greater uncertainty than the other methods in estimating the interaction effect (Figure 6).

5.2. Results for Artificial Datasets. Tables 1–6 show RB, PB, CR, model-based SE, empirical SE, and λ under the missing mechanism of MCAR and MAR for imputation methods; the tables are separated by different measurements.

By comparing the values of RB for each coefficient, interesting results were obtained: for all scenarios, the MICE-Interaction method performed very well in terms of RB since all RB values for this method were almost zero. For instance, at 40% missing percentage under both mechanisms, the amount of RBs for the interaction produced by the MICE-Interaction method was equal to zero (Table 1).

Considering the PB values, the MICE-Interaction method had acceptable performance since the PB values were less than 5%. For all the coefficients which were in the interaction effect, at all missing percentages in both missing mechanisms, the MICE-Interaction method led to a negligible PB (PB < 2.2%) and had relatively good performance (Table 2).

MICE-Interaction, which correctly includes the interaction term in the imputation model, led to slightly higher coverage for the interaction effect than the other imputation methods. The CR values of the MICE-Interaction method were more than 95%. In almost all scenarios, the CR values of MICE-Stratified method were less than 95% (Table 3).

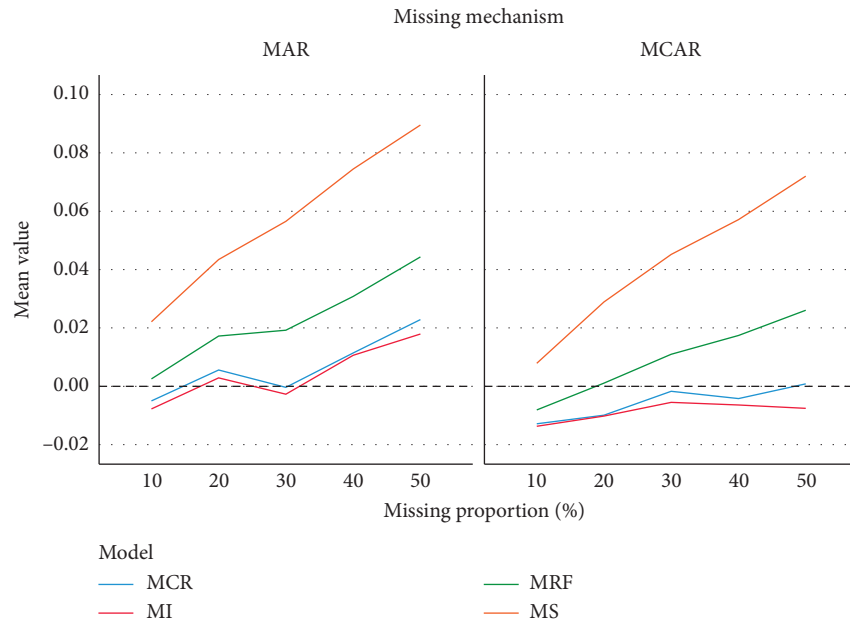


FIGURE 1: Comparison of the performance of imputation methods for real dataset based on the average raw bias (RB) for the interaction effect. MAR, missing at random; MCAR, missing completely at random; MCR, multiple imputation by chained equation using classification and regression trees; MRF, multiple imputation by chained equation using random forests; MI, MICE-Interaction method; MS, MICE-Stratified method.

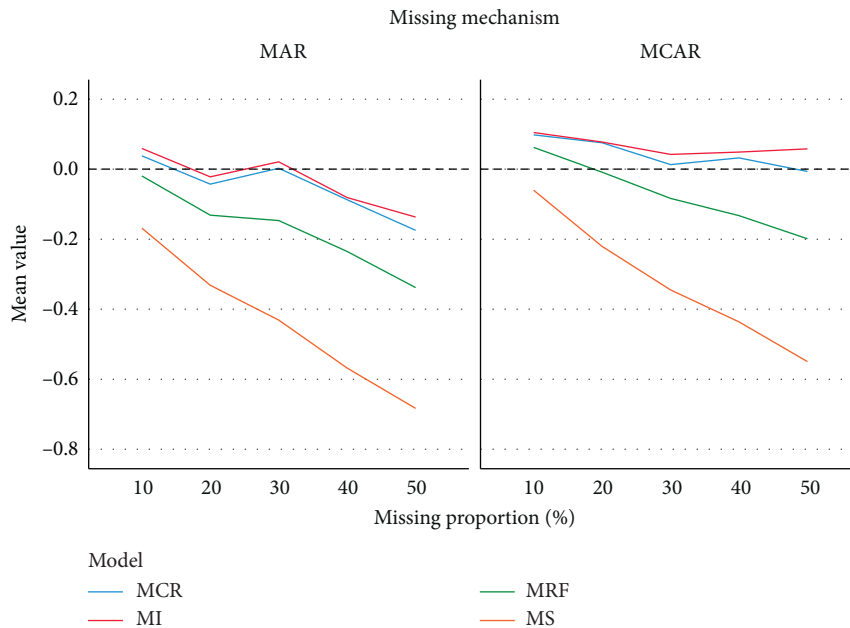


FIGURE 2: Comparison of the performance of imputation methods for real dataset based on the average percent bias (PB) for the interaction effect. MAR, missing at random; MCAR, missing completely at random; MCR, multiple imputation by chained equation using classification and regression trees; MRF, multiple imputation by chained equation using random forests; MI, MICE-Interaction method; MS, MICE-Stratified method.

By examining the model and empirical SEs, it is determined that by simulations the difference between the empirical and Rubin’s MI variance was small or negligible. The model SEs was larger than the empirical SEs, and this allowed the coverage to be relatively good despite the slight bias (Tables 4-5).

5.2.1. *The Proportion of Variation Attributable to the Missing Data (λ)*. Examining the estimated values of λ , it is interesting to note that, in all scenarios, all methods had acceptable performance based on the λ criterion ($\lambda < 0.5$). The λ values obtained from the MICE-CART and MICE-Stratified are lower than other methods. The MICE-Stratified

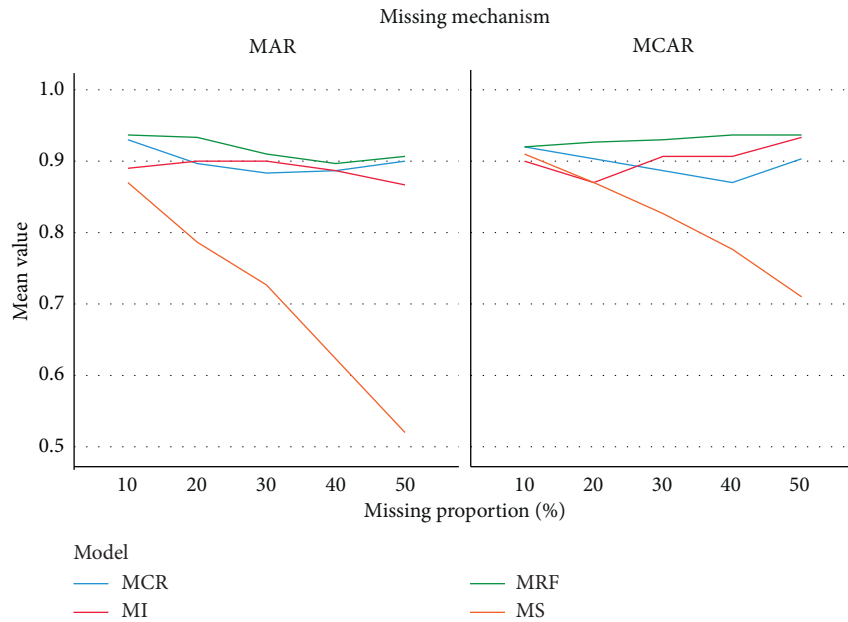


FIGURE 3: Comparison of the performance of imputation methods for real dataset based on the average coverage rate (CR) for the interaction effect. MAR, missing at random; MCAR, missing completely at random; MCR, multiple imputation by chained equation using classification and regression trees; MRF, multiple imputation by chained equation using random forests; MI, MICE-Interaction method; MS, MICE-Stratified method.

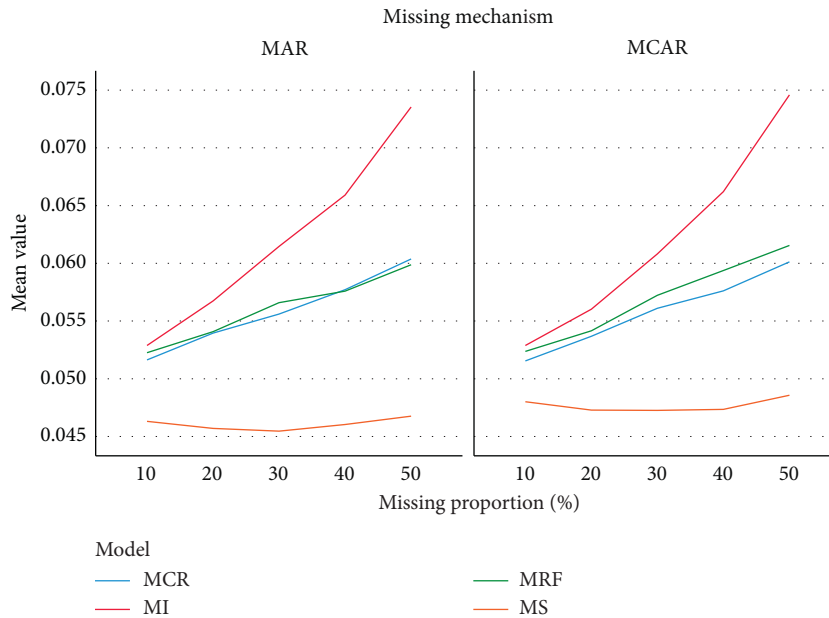


FIGURE 4: Comparison of the performance of imputation methods for real dataset based on the model SE for the interaction effect. MAR, missing at random; MCAR, missing completely at random; MCR, multiple imputation by chained equation using classification and regression trees; MRF, multiple imputation by chained equation using random forests; MI, MICE-Interaction method; MS, MICE-Stratified method.

method resulted in the lowest λ which showed that the influence of the imputation model on the final result was not larger than that of the complete-data model. The estimated values were less than 0.2, indicating that the methods deal with a relatively large fraction of missing information (Table 6).

6. Discussion

In this study, four MI methods were evaluated, including MICE-Stratified, MICE-Interaction, and two recursive partitioning techniques that incorporate interaction effects in the data, as imputation methods in MICE: CART and RF.

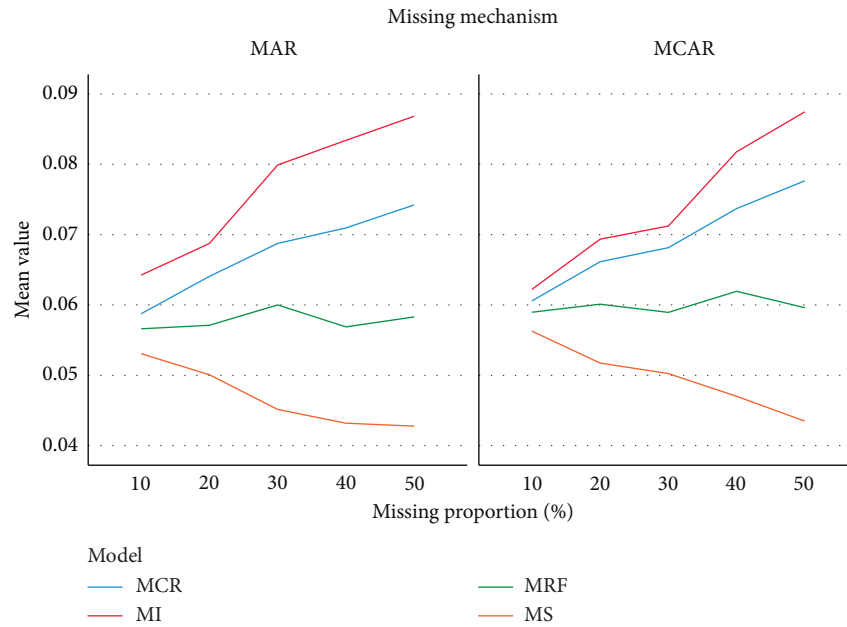


FIGURE 5: Comparison of the performance of imputation methods for real dataset based on the empirical SE. MAR, missing at random; MCAR, missing completely at random; MCR, multiple imputation by chained equation using classification and regression trees; MRF, multiple imputation by chained equation using random forests; MI, MICE-Interaction method; MS, MICE-Stratified method.

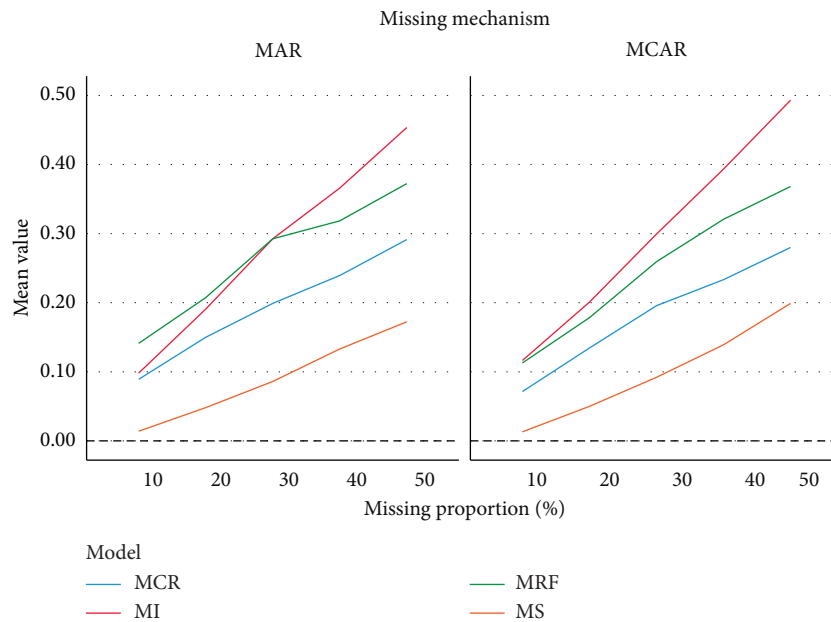


FIGURE 6: Comparison of the performance of imputation methods for real dataset based on the proportion of the variation attributable to the missing data (λ) for the interaction effect.

In addition, the percentage of missing data and two missing mechanisms on the performance of the MI methods were examined. We studied the raw bias, percent bias, coverage rates, model-based SE, empirical SE, and the proportion of variation attributable to the missing data, after imputation by these methods. Overall, at lower percentages of missing values (at most 30%), all three methods except MICE-Stratified appeared to produce reasonably high quality imputations.

Consistent with similar studies, MI methods were evaluated based on RB, PB, CR, model SE, empirical SE, and λ [3, 22]. RB, PB, and CR are several measures that may inform us about the statistical validity or precision of a particular procedure. RB is a measure of invalidity and should be close zero. For acceptable performance, we used an upper limit for PB of 5% [3, 32]. CR is also a measure of validity but at the same time is a function of the average width of the confidence intervals, which is a measure of

TABLE 1: Comparison of the performance of imputation methods based on the average raw bias (RB) of regression coefficients over 1000 simulations.

Variable	Method	MAR					MCAR				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
X_1 (BMReal)	MCR	0.024	0.050	0.082	0.119	0.145	0.028	0.054	0.083	0.112	0.157
	MI	0.001	0.006	0.012	0.016	0.020	0.002	0.008	0.012	0.000	0.020
	MRF	-0.013	-0.024	-0.028	-0.038	-0.052	-0.009	-0.017	-0.026	-0.040	-0.040
	MS	-0.071	-0.123	-0.177	-0.228	-0.278	-0.069	-0.129	-0.178	-0.236	-0.276
X_2 (Betotal)	MCR	-0.010	-0.021	-0.034	-0.050	-0.058	-0.011	-0.023	-0.032	-0.044	-0.066
	MI	0.001	0.001	0.004	0.000	0.009	0.000	0.000	0.004	0.008	0.003
	MRF	-0.034	-0.064	-0.092	-0.124	-0.142	-0.031	-0.062	-0.087	-0.112	-0.139
	MS	-0.087	-0.156	-0.213	-0.266	-0.308	-0.086	-0.157	-0.218	-0.267	-0.314
X_3 (PACS4)	MCR	-0.023	-0.036	-0.052	-0.073	-0.086	-0.010	-0.016	-0.030	-0.041	-0.053
	MI	-0.003	0.004	0.006	0.008	0.011	0.001	0.005	0.005	0.008	0.008
	MRF	-0.054	-0.091	-0.129	-0.167	-0.196	-0.034	-0.066	-0.098	-0.123	-0.151
	MS	-0.118	-0.184	-0.236	-0.278	-0.311	-0.086	-0.156	-0.217	-0.264	-0.312
X_4 (PSPS)	MCR	-0.020	-0.036	-0.055	-0.071	-0.095	-0.008	-0.022	-0.034	-0.046	-0.063
	MI	0.000	-0.001	0.002	-0.001	0.000	0.003	0.002	0.002	0.005	0.004
	MRF	-0.052	-0.094	-0.129	-0.167	-0.200	-0.032	-0.066	-0.097	-0.126	-0.156
	MS	-0.113	-0.187	-0.237	-0.280	-0.314	-0.087	-0.158	-0.218	-0.271	-0.315
X_5 (Gender)	MCR	0.000	-0.005	-0.022	-0.033	-0.035	-0.012	-0.025	-0.040	-0.062	-0.068
	MI	0.004	0.007	0.003	0.003	0.018	-0.001	0.004	-0.006	-0.002	0.003
	MRF	-0.015	-0.027	-0.054	-0.071	-0.080	-0.024	-0.043	-0.068	-0.086	-0.102
	MS	-0.043	-0.070	-0.097	-0.121	-0.132	-0.051	-0.084	-0.115	-0.141	-0.162
$X_1 * X_5$ (BMReal * Gender)	MCR	-0.076	-0.144	-0.228	-0.304	-0.390	-0.072	-0.136	-0.225	-0.300	-0.401
	MI	-0.002	0.000	-0.010	0.006	-0.003	0.002	0.001	0.000	0.003	-0.004
	MRF	-0.128	-0.233	-0.345	-0.431	-0.517	-0.119	-0.231	-0.327	-0.417	-0.514
	MS	-0.236	-0.405	-0.533	-0.628	-0.716	-0.231	-0.399	-0.534	-0.631	-0.722

TABLE 2: Comparison of the performance of imputation methods based on the average percent bias (PB) of regression coefficients over 1000 simulations.

Variable	Method	MAR					MCAR				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
X_1 (BMReal)	MCR	0.048	0.101	0.165	0.239	0.291	0.056	0.109	0.167	0.226	0.315
	MI	0.002	0.012	0.025	0.033	0.040	0.005	0.016	0.024	0.000	0.039
	MRF	-0.025	-0.048	-0.056	-0.077	-0.103	-0.018	-0.034	-0.052	-0.079	-0.079
	MS	-0.143	-0.246	-0.355	-0.455	-0.556	-0.138	-0.259	-0.355	-0.471	-0.552
X_2 (Betotal)	MCR	-0.020	-0.043	-0.068	-0.101	-0.117	-0.021	-0.047	-0.064	-0.089	-0.132
	MI	0.002	0.001	0.008	0.000	0.016	0.001	0.000	0.009	0.016	0.007
	MRF	-0.068	-0.129	-0.185	-0.248	-0.286	-0.063	-0.124	-0.175	-0.224	-0.279
	MS	-0.176	-0.314	-0.428	-0.533	-0.620	-0.173	-0.316	-0.437	-0.535	-0.631
X_3 (PACS4)	MCR	-0.047	-0.072	-0.103	-0.146	-0.172	-0.020	-0.032	-0.061	-0.081	-0.106
	MI	-0.006	0.008	0.012	0.016	0.022	0.001	0.009	0.010	0.016	0.016
	MRF	-0.109	-0.181	-0.258	-0.334	-0.393	-0.069	-0.132	-0.196	-0.247	-0.302
	MS	-0.235	-0.368	-0.472	-0.556	-0.622	-0.172	-0.312	-0.433	-0.527	-0.623
X_4 (PSPS)	MCR	-0.041	-0.072	-0.111	-0.144	-0.190	-0.017	-0.044	-0.068	-0.093	-0.126
	MI	0.000	-0.003	0.005	-0.001	0.000	0.005	0.004	0.003	0.009	0.010
	MRF	-0.104	-0.188	-0.259	-0.335	-0.401	-0.065	-0.134	-0.196	-0.254	-0.312
	MS	-0.227	-0.375	-0.477	-0.563	-0.631	-0.175	-0.318	-0.438	-0.545	-0.633
X_5 (Gender)	MCR	0.001	-0.021	-0.088	-0.129	-0.137	-0.047	-0.099	-0.161	-0.247	-0.279
	MI	0.013	0.027	0.008	0.012	0.068	-0.005	0.014	-0.027	-0.010	0.008
	MRF	-0.060	-0.109	-0.214	-0.283	-0.320	-0.093	-0.172	-0.273	-0.341	-0.408
	MS	-0.172	-0.279	-0.384	-0.479	-0.522	-0.202	-0.338	-0.458	-0.563	-0.646
$X_1 * X_5$ (BMReal * Gender)	MCR	-0.076	-0.144	-0.228	-0.304	-0.391	-0.072	-0.136	-0.225	-0.301	-0.401
	MI	-0.002	0.000	-0.010	0.006	-0.002	0.002	0.001	-0.001	0.003	-0.003
	MRF	-0.128	-0.233	-0.345	-0.431	-0.518	-0.119	-0.231	-0.328	-0.417	-0.514
	MS	-0.236	-0.406	-0.533	-0.629	-0.716	-0.231	-0.399	-0.534	-0.631	-0.722

TABLE 3: Comparison of the performance of imputation methods based on the CRs of regression coefficients over 1000 simulations.

Variable	Method	MAR					MCAR				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
X_1 (BMlreal)	MCR	1.000	1.000	0.981	0.964	0.885	1.000	1.000	0.982	0.955	0.924
	MI	1.000	1.000	0.998	0.997	0.984	1.000	1.000	0.998	0.992	0.985
	MRF	1.000	1.000	1.000	0.998	0.986	1.000	1.000	1.000	0.998	0.994
	MS	0.996	0.967	0.826	0.702	0.556	0.997	0.953	0.850	0.674	0.563
X_2 (Betotal)	MCR	1.000	0.999	0.982	0.944	0.906	1.000	0.997	0.987	0.955	0.901
	MI	1.000	0.999	0.998	0.994	0.983	1.000	1.000	0.998	0.995	0.986
	MRF	1.000	0.999	0.987	0.942	0.903	1.000	1.000	0.991	0.963	0.875
	MS	0.983	0.740	0.415	0.205	0.092	0.973	0.727	0.402	0.208	0.078
X_3 (PACS4)	MCR	1.000	0.996	0.965	0.893	0.834	1.000	0.997	0.987	0.961	0.907
	MI	1.000	1.000	0.998	0.991	0.987	1.000	0.999	0.999	0.998	0.983
	MRF	1.000	0.980	0.919	0.798	0.678	1.000	0.997	0.964	0.926	0.850
	MS	0.896	0.500	0.210	0.093	0.049	0.973	0.666	0.301	0.140	0.046
X_4 (PSPS)	MCR	1.000	0.994	0.955	0.907	0.824	1.000	1.000	0.983	0.946	0.899
	MI	1.000	1.000	0.997	0.995	0.985	1.000	1.000	1.000	0.998	0.989
	MRF	1.000	0.980	0.920	0.789	0.676	1.000	0.999	0.970	0.904	0.830
	MS	0.893	0.489	0.205	0.088	0.042	0.978	0.648	0.295	0.110	0.047
X_5 (Gender)	MCR	1.000	0.999	0.995	0.981	0.959	1.000	0.999	0.990	0.979	0.957
	MI	1.000	1.000	0.998	0.997	0.989	1.000	1.000	0.999	0.996	0.988
	MRF	1.000	1.000	0.998	0.999	0.994	1.000	1.000	0.999	0.997	0.992
	MS	1.000	0.991	0.974	0.951	0.962	0.999	0.991	0.974	0.958	0.924
$X_1 * X_5$ (BMlreal *Gender)	MCR	1.000	0.997	0.929	0.822	0.712	1.000	0.996	0.946	0.834	0.677
	MI	1.000	1.000	0.999	0.998	0.990	1.000	0.999	0.999	0.987	0.983
	MRF	1.000	0.988	0.867	0.714	0.558	1.000	0.984	0.908	0.734	0.545
	MS	0.910	0.455	0.173	0.072	0.015	0.916	0.469	0.166	0.037	0.014

TABLE 4: Comparison of the performance of imputation methods based on the model based SE of regression coefficients over 1000 simulations.

Variable	Method	MAR					MCAR				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
X_1 (BMlreal)	MCR	0.161	0.167	0.174	0.183	0.190	0.161	0.169	0.174	0.182	0.190
	MI	0.164	0.176	0.190	0.209	0.233	0.165	0.177	0.191	0.208	0.230
	MRF	0.163	0.171	0.178	0.185	0.189	0.164	0.171	0.178	0.184	0.192
	MS	0.149	0.146	0.146	0.146	0.149	0.149	0.146	0.146	0.146	0.148
X_2 (Betotal)	MCR	0.114	0.118	0.122	0.124	0.128	0.115	0.119	0.122	0.125	0.128
	MI	0.117	0.125	0.135	0.147	0.163	0.118	0.126	0.136	0.148	0.163
	MRF	0.116	0.120	0.125	0.129	0.133	0.117	0.121	0.125	0.128	0.131
	MS	0.105	0.102	0.101	0.101	0.102	0.105	0.102	0.101	0.101	0.102
X_3 (PACS4)	MCR	0.105	0.109	0.112	0.113	0.117	0.105	0.109	0.112	0.114	0.118
	MI	0.108	0.118	0.126	0.140	0.154	0.108	0.115	0.125	0.136	0.147
	MRF	0.107	0.111	0.114	0.117	0.119	0.106	0.111	0.114	0.118	0.120
	MS	0.095	0.093	0.092	0.092	0.093	0.096	0.093	0.092	0.092	0.093
X_4 (PSPS)	MCR	0.105	0.109	0.112	0.114	0.116	0.105	0.108	0.111	0.114	0.117
	MI	0.109	0.117	0.127	0.140	0.154	0.108	0.115	0.124	0.135	0.149
	MRF	0.107	0.111	0.115	0.117	0.119	0.106	0.110	0.113	0.117	0.120
	MS	0.095	0.093	0.092	0.092	0.093	0.096	0.093	0.092	0.092	0.093
X_5 (Gender)	MCR	0.192	0.198	0.204	0.210	0.215	0.192	0.199	0.204	0.210	0.215
	MI	0.197	0.210	0.228	0.245	0.270	0.198	0.212	0.228	0.249	0.274
	MRF	0.194	0.202	0.208	0.216	0.219	0.195	0.204	0.209	0.217	0.222
	MS	0.178	0.175	0.174	0.175	0.178	0.179	0.175	0.174	0.176	0.178
$X_1 * X_5$ (BMlreal *Gender)	MCR	0.238	0.246	0.254	0.264	0.270	0.239	0.249	0.255	0.261	0.268
	MI	0.243	0.261	0.281	0.307	0.339	0.243	0.261	0.280	0.305	0.342
	MRF	0.241	0.248	0.253	0.259	0.260	0.241	0.249	0.254	0.258	0.263
	MS	0.209	0.199	0.192	0.189	0.189	0.210	0.199	0.193	0.189	0.188

TABLE 5: Comparison of the performance of imputation methods based on the empirical SE of regression coefficients over 1000 simulations.

Variable	Method	MAR					MCAR				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
X_1 (BMIreal)	MCR	0.069	0.102	0.135	0.160	0.207	0.072	0.102	0.127	0.155	0.181
	MI	0.063	0.089	0.122	0.141	0.187	0.065	0.093	0.117	0.149	0.172
	MRF	0.060	0.082	0.101	0.110	0.124	0.064	0.083	0.095	0.111	0.122
	MS	0.079	0.094	0.109	0.110	0.113	0.078	0.094	0.100	0.107	0.110
X_2 (Betotal)	MCR	0.055	0.077	0.100	0.120	0.146	0.054	0.073	0.093	0.114	0.136
	MI	0.050	0.067	0.088	0.103	0.130	0.050	0.068	0.083	0.105	0.132
	MRF	0.050	0.062	0.072	0.080	0.093	0.049	0.061	0.071	0.082	0.092
	MS	0.058	0.068	0.073	0.078	0.079	0.058	0.068	0.072	0.078	0.076
X_3 (PACS4)	MCR	0.055	0.074	0.093	0.116	0.140	0.053	0.071	0.088	0.106	0.127
	MI	0.051	0.066	0.083	0.105	0.122	0.048	0.067	0.079	0.095	0.121
	MRF	0.051	0.060	0.069	0.079	0.087	0.047	0.060	0.067	0.074	0.083
	MS	0.059	0.067	0.069	0.074	0.070	0.056	0.063	0.069	0.071	0.069
X_4 (PSPS)	MCR	0.055	0.077	0.094	0.113	0.141	0.051	0.070	0.087	0.110	0.126
	MI	0.048	0.068	0.082	0.102	0.122	0.046	0.064	0.079	0.098	0.118
	MRF	0.049	0.061	0.069	0.079	0.087	0.045	0.058	0.066	0.077	0.081
	MS	0.059	0.066	0.067	0.071	0.069	0.054	0.065	0.069	0.071	0.069
X_5 (Gender)	MCR	0.077	0.113	0.145	0.172	0.194	0.083	0.114	0.152	0.172	0.205
	MI	0.072	0.105	0.136	0.168	0.208	0.073	0.108	0.143	0.171	0.217
	MRF	0.070	0.095	0.111	0.122	0.135	0.071	0.094	0.118	0.120	0.135
	MS	0.089	0.111	0.125	0.134	0.126	0.091	0.114	0.122	0.121	0.134
$X_1 * X_5$ (BMIreal *Gender)	MCR	0.096	0.139	0.182	0.216	0.248	0.102	0.143	0.175	0.217	0.242
	MI	0.086	0.126	0.172	0.214	0.265	0.089	0.131	0.173	0.219	0.258
	MRF	0.085	0.115	0.132	0.143	0.147	0.088	0.116	0.131	0.143	0.143
	MS	0.127	0.144	0.156	0.151	0.149	0.123	0.142	0.148	0.142	0.143

TABLE 6: Comparison of the performance of imputation methods based on the proportion of the variation attributable to the missing data (λ) of regression coefficients over 1000 simulations.

Variable	Method	MAR					MCAR				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
X_1 (BMIreal)	MCR	0.069	0.121	0.178	0.232	0.267	0.067	0.132	0.171	0.222	0.264
	MI	0.106	0.206	0.299	0.398	0.489	0.109	0.213	0.308	0.395	0.485
	MRF	0.116	0.206	0.270	0.330	0.364	0.119	0.199	0.265	0.317	0.373
	MS	0.010	0.038	0.080	0.123	0.173	0.010	0.038	0.076	0.121	0.171
X_2 (Betotal)	MCR	0.076	0.139	0.192	0.233	0.272	0.078	0.143	0.196	0.232	0.271
	MI	0.111	0.205	0.296	0.394	0.486	0.113	0.213	0.309	0.394	0.484
	MRF	0.125	0.213	0.290	0.345	0.392	0.133	0.217	0.279	0.332	0.369
	MS	0.012	0.040	0.079	0.125	0.173	0.011	0.041	0.079	0.121	0.169
X_3 (PACS4)	MCR	0.093	0.156	0.205	0.233	0.279	0.080	0.143	0.200	0.230	0.279
	MI	0.127	0.243	0.321	0.426	0.510	0.112	0.210	0.308	0.401	0.472
	MRF	0.149	0.238	0.299	0.351	0.382	0.128	0.219	0.284	0.341	0.381
	MS	0.014	0.046	0.084	0.123	0.175	0.012	0.038	0.079	0.129	0.172
X_4 (PSPS)	MCR	0.091	0.157	0.209	0.243	0.273	0.080	0.143	0.191	0.231	0.271
	MI	0.130	0.239	0.329	0.429	0.508	0.113	0.216	0.307	0.395	0.486
	MRF	0.152	0.236	0.304	0.351	0.384	0.130	0.218	0.274	0.327	0.380
	MS	0.014	0.046	0.084	0.129	0.174	0.012	0.041	0.079	0.124	0.173
X_5 (Gender)	MCR	0.072	0.129	0.180	0.228	0.264	0.072	0.132	0.175	0.223	0.258
	MI	0.105	0.201	0.304	0.384	0.472	0.112	0.217	0.306	0.400	0.482
	MRF	0.116	0.199	0.261	0.323	0.349	0.114	0.208	0.259	0.319	0.358
	MS	0.010	0.038	0.076	0.119	0.171	0.011	0.038	0.076	0.124	0.170
$X_1 * X_5$ (BMIreal *Gender)	MCR	0.095	0.159	0.219	0.274	0.309	0.095	0.171	0.220	0.263	0.301
	MI	0.111	0.213	0.307	0.398	0.483	0.107	0.214	0.300	0.391	0.491
	MRF	0.160	0.250	0.311	0.368	0.404	0.159	0.249	0.311	0.356	0.408
	MS	0.015	0.048	0.089	0.135	0.187	0.015	0.047	0.092	0.134	0.185

precision. If all is well, then RB should be close to zero, and the coverage should be near 0.95. Methods having no bias and proper coverage are called randomization-valid [33]. If two methods are both randomization-valid, the method with the shorter confidence intervals is more efficient. The λ value is a function of the number of imputations and interpreted as the proportion of the variation attributable to the missing data. If λ is high, say $\lambda > 0.5$, the influence of the imputation model on the final result is larger than that of the complete-data model. Better imputation results in lower values of λ [2, 34, 35]. As we saw earlier in the current study, in almost all scenarios, the λ values were less than 0.5. Although the lowest amount of λ was attributed to the MICE-Stratified and MICE-CART method, according to the RB and CR values of these two methods, the MICE-CART method was superior to the MICE-Stratified method.

Our main result is that, relative to one another, the MICE-Interaction method appeared to have the best performance at all missing percentages. Also, the recursive partitioning methods had better performance than MICE-Stratified method. Furthermore, the simulation results of comparing the methods showed that the quality of parameter estimates was almost identical in the two missing mechanisms.

An extension of this study would consider a dataset that has more complex structures between the variables than our data. In addition, looking at each model closely, one would observe that the design of this study directly aligns with the strength of each recursive partitioning model. The recursive splits in the tree structure allow CART to effectively capture higher order interaction more accurately, as long as the number of individuals in each leaf is large enough.

The imperfect imputation models that resulted in the bias of recursive partitioning techniques may have emerged from the presence of main effects in the data. That is, recursive partitioning techniques have difficulty in modelling linear main effects. It is hard to capture the main effects because, due to the binary tree model, it would take many fortuitous splits to recreate the structure [36]. We expect that this problem will also happen implementing other recursive partitioning techniques. The solution to this problem is provided by STIMA [37], which combines a linear main effects model with recursive partitioning. The difficulty with modelling linear main effects also explains why in our study the recursive partitioning imputation methods led to biases that are somewhat higher for the main effects compared to MICE-Interaction method.

Meng stated that the analysis procedure should be congenial to the imputation model. Bartlett et al. connected congeniality to compatibility by extending the joint distribution of the imputation model to include the substantive model [38, 39]. Uncongeniality can occur if the imputation model is specified as more restrictive than the complete-data model, or if it fails to account for important factors in the missing data mechanism. Both types of missions introduce biased and possibly inefficient estimates [3]. In the MICE-Interaction method, we used the imputation model, which included the interaction effect, and this made compatibility between imputation and analysis model [40]. Therefore, the

method led to less bias in estimating the coefficients. The higher biases for the interaction effects by RF compared to CART may be explained by interactions that are missed in the tree building process due to drawing bootstrap samples and the (low) number of randomly preselected variables [41]. Therefore, it is concluded that MICE-Interaction preserved the interaction effect best and the MICE-Interaction method is recommended if a user has presumptions of interaction effects. The quality of imputations for any of the methods was lower for datasets with higher percentage of missing data.

Some features of the present study could be considered as strengths. For example, to our knowledge, this is the first time that the performance of four imputation methods was evaluated in observations involving a combination of binary and continuous predictors, in the presence of an interaction effect.

To improve on the performance of the MICE-RF, an extension of this study will look into the possibility of changing the number of trees in RF algorithm. Future work will consider the missing not at random mechanism instead of missing completely at random and missing at random considered here. Finally, as mentioned earlier, datasets with more complex structures between the variables will be considered.

7. Conclusion

In summary, this paper offered a fair comparison between MICE-stratified, MICE-Interaction, and tree-based imputation methods in the MICE algorithm. To our knowledge, this is the first paper to compare tree-based imputation in MICE to a parametric model that includes a true interaction effect between a dummy and a continuous variable.

MICE-Interaction had the highest coverage of the interaction effect. Importantly, parametric imputation should only be utilized if there is enough information to ensure that all necessary interaction terms are included in the imputation model. If one can accept the reduction of coverage for the interaction effect, recursive partitioning methods are recommended as they do not require specification of the imputation model.

Beyond the fact that there is still room for improvement, it can be concluded that although recursive partitioning methods were valuable for imputing datasets containing interaction effect between a binary and a continuous variable, properly incorporating interactions in the parametric imputation model (MICE-Interaction method) led to much better performance.

Abbreviations

CR:	Coverage rate
FCS:	Full conditional specification
JAV:	Just another variable
MCR:	MICE-CART
MRF:	MICE-RF
MI:	MICE-Interaction
MS:	MICE-Stratified
MAR:	Missing at random

MCAR: Missing completely at random
 MI: Multiple imputation
 PB: Percent bias
 PMM: Predictive mean matching
 RB: Raw bias
 RF: Random forests
 SE: Standard error.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request to the corresponding author. The corresponding author had previously used the real data from Addressing the Problem of Missing Data in Decision Tree Modelling (link: <https://doi.org/10.1080/02664763.2017.1284184>).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] J. A. C. Sterne, I. R. White, J. B. Carlin et al., "Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls," *BMJ*, vol. 338, no. 1, Article ID b2393, 2009.
- [2] D. B. Rubin, *Multiple Imputation for Nonresponse in Surveys*, John Wiley & Sons, Hoboken, NJ, USA, 2004.
- [3] S. Van Buuren, *Flexible Imputation of Missing Data*, CRC Press, Boca Raton, FL, USA, 2018.
- [4] D. B. Rubin, "Multiple imputation after 18+ years," *Journal of the American Statistical Association*, vol. 91, no. 434, pp. 473–489, 1996.
- [5] J. Barnard and X.-L. Meng, "Applications of multiple imputation in medical studies: from AIDS to NHANES," *Statistical Methods in Medical Research*, vol. 8, no. 1, pp. 17–36, 1999.
- [6] R. J. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, John Wiley & Sons, Hoboken, NJ, USA, 2019.
- [7] S. Van Buuren and K. Oudshoorn, *Flexible Multivariate Imputation by MICE*, TNO, Leiden, Netherlands, 1999.
- [8] S. Van Buuren, "Multiple imputation of discrete and continuous data by fully conditional specification," *Statistical Methods in Medical Research*, vol. 16, no. 3, pp. 219–242, 2007.
- [9] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer Science & Business Media, Berlin, Germany, 2008.
- [10] F. Li, M. Baccini, F. Mealli, E. R. Zell, C. E. Frangakis, and D. B. Rubin, "Multiple imputation by ordered monotone blocks with application to the anthrax vaccine research program," *Journal of Computational and Graphical Statistics*, vol. 23, no. 3, pp. 877–892, 2014.
- [11] T. E. Raghunathan and D. B. Rubin, "Roles for Bayesian techniques in survey sampling," in *Proceedings of the Silver Jubilee Meeting of the Statistical Society of Canada*, Canada, 1998.
- [12] S. Buuren and K. Groothuis-Oudshoorn, "MICE: multivariate imputation by chained equations in R," *Journal of Statistical Software*, vol. 45, no. 3, pp. 1–68, 2010.
- [13] S. Yang, *Flexible Imputation of Missing Data*, Chapman & Hall/CRC Press, Boca Raton, FL, USA, 2018.
- [14] Y.-S. Su, A. E. Gelman, J. Hill, and M. Yajima, "Multiple imputation with diagnostics (Mi) in R: opening windows into the black box," *Journal of Statistical Software*, vol. 45, no. 2, pp. 1–31, 2011.
- [15] P. Royston and I. R. White, "Multiple imputation by chained equations (MICE): implementation in Stata," *Journal of Statistical Software*, vol. 45, no. 4, pp. 1–20, 2011.
- [16] S. R. Seaman, J. W. Bartlett, and I. R. White, "Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods," *BMC Medical Research Methodology*, vol. 12, no. 1, p. 46, 2012.
- [17] J. N. Morgan and J. A. Sonquist, "Problems in the analysis of survey data, and a proposal," *Journal of the American Statistical Association*, vol. 58, no. 302, pp. 415–434, 1963.
- [18] L. F. Burgette and J. P. Reiter, "Multiple imputation for missing data via sequential regression trees," *American Journal of Epidemiology*, vol. 172, no. 9, pp. 1070–1076, 2010.
- [19] J. L. Schafer, *Analysis of Incomplete Multivariate Data*, Chapman and Hall/CRC, Boca Raton, FL, USA, 1997.
- [20] L. L. Doove, S. Van Buuren, and E. Dusseldorp, "Recursive partitioning for missing data imputation in the presence of interaction effects," *Computational Statistics & Data Analysis*, vol. 72, pp. 92–104, 2014.
- [21] A. D. Shah, J. W. Bartlett, J. Carpenter, O. Nicholas, and H. Hemingway, "Comparison of random forest and parametric imputation models for imputing missing data using MICE: a CALIBER study," *American Journal of Epidemiology*, vol. 179, no. 6, pp. 764–774, 2014.
- [22] T. Therneau, B. Atkinson, and B. Ripley, *Recursive Partitioning and Regression Trees. R Package 'rpart' (Version 4.1-11)*, R Found Statistical Computing, Vienna, Austria, 2017.
- [23] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
- [24] C. Strobl, A.-L. Boulesteix, and T. Augustin, "Unbiased split selection for classification trees based on the Gini index," *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 483–501, 2007.
- [25] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*, Springer Series in Statistics, New York, NY, USA, 2001.
- [26] L. F. J. Breiman and R. A. Olshen, *Classification and Regression Trees*, Chapman and Hall/CRC, Boca Raton, FL, USA, 1984.
- [27] B. Garrusi, S. Garousi, and M. R. Baneshi, "Body image and body change: predictive factors in an Iranian population," *International Journal of Preventive Medicine*, vol. 4, no. 8, pp. 940–948, 2013.
- [28] J. L. Schafer, "Multiple imputation: a primer," *Statistical Methods in Medical Research*, vol. 8, no. 1, pp. 3–15, 1999.
- [29] H. Demirtas, S. A. Freels, and R. M. Yucel, "Plausibility of multivariate normality assumption when multiply imputing non-Gaussian continuous outcomes: a simulation assessment," *Journal of Statistical Computation and Simulation*, vol. 78, no. 1, pp. 69–84, 2008.
- [30] H. Demirtas, "Simulation driven inferences for multiply imputed longitudinal datasets," *Statistica Neerlandica*, vol. 58, no. 4, pp. 466–482, 2004.
- [31] L. M. Collins, J. L. Schafer, and C.-M. Kam, "A comparison of inclusive and restrictive strategies in modern missing data procedures," *Psychological Methods*, vol. 6, no. 4, pp. 330–351, 2001.
- [32] H. Demirtas and D. Hedeker, "Multiple imputation under power polynomials," *Communications in Statistics—Simulation and Computation*, vol. 37, no. 8, pp. 1682–1695, 2008.

- [33] D. Rubin, *Multiple Imputation for Nonresponse in Surveys*, John Wiley & Sons, New York, NY, USA, 1987.
- [34] C. A. Bernaards, M. M. Farmer, K. Qi, G. S. Dulai, P. A. Ganz, and K. L. Kahn, "Comparison of two multiple imputation procedures in a cancer screening survey," 2002.
- [35] StataCorp LLC, *Stata Multiple-Imputation Reference Manual*, StataCorp LLC, College Station, TX, USA, 2013.
- [36] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Science & Business Media, Berlin, Germany, 2009.
- [37] E. Dusseldorp, C. Conversano, and B. J. Van Os, "Combining an additive and tree-based regression model simultaneously: STIMA," *Journal of Computational and Graphical Statistics*, vol. 19, no. 3, pp. 514–530, 2010.
- [38] X.-L. Meng, "Multiple-imputation inferences with uncongenial sources of input," *Statistical Science*, vol. 9, no. 4, pp. 538–558, 1994.
- [39] J. W. Bartlett, S. R. Seaman, I. R. White, and J. R. Carpenter, "Multiple imputation of covariates by fully conditional specification: accommodating the substantive model," *Statistical Methods in Medical Research*, vol. 24, no. 4, pp. 462–487, 2015.
- [40] E. Slade and M. G. Naylor, "A fair comparison of tree-based and parametric methods in multiple imputation by chained equations," *Statistics in Medicine*, vol. 39, no. 8, pp. 1156–1166, 2020.
- [41] C. Strobl, J. Malley, and G. Tutz, "An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests," *Psychological Methods*, vol. 14, no. 4, pp. 323–348, 2009.