Comparison of Performance of Recursive Partitioning Models and Parametric Imputation Models Based On Predictive Mean Matching In Multiple Imputation By Chained Equations, in Order to Impute The Missing Values of The Binary Outcome, in The Presence of An Interaction Effect

Sara Javadi

Kerman University of Medical Sciences

Abbas Bahrampour

Kerman University of Medical Sciences

Mohammad Mehdi Saber

Higher Education Center of Eghlid

Mohammad Reza Baneshi (✉ rbaneshi2@gmail.com)

Kerman University of Medical Sciences

Research Article

Keywords: Imputation, Interaction, Missing data, Regression tree, Predictive Mean Matching, Recursive Partitioning

Posted Date: October 28th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-961665/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Comparison of performance of recursive partitioning models and parametric imputation models based on predictive mean matching in multiple imputation by chained equations, in order to impute the missing values of the binary outcome, in the presence of an interaction effect

Sara Javadi¹, Abbas Bahrampour², Mohammad Mehdi Saber³, Mohammad Reza Baneshi²

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

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

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

***Corresponding author:** Mohammad Reza Baneshi; E-mail: m_baneshi@kmu.ac.ir & rbaneshi2@gmail.com.

Declaration section

- Ethics approval and consent to participate (Not applicable)
- Consent for publication (Not applicable)
- Availability of data and materials (Not applicable)
- Competing interests

We declare that the authors have no competing interests as defined by BMC, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

- Funding

This manuscript is taken from a doctoral dissertation. It has no financial support. We are sending a manuscript from Iran, which is poor and low-income, we request that the paper be published free of charge.

- Authors' contributions

Sara Javadi: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Writing-original draft.

Abbas Bahrapour: Formal analysis, Investigation, Methodology, Software, Writing-original draft.

Mohammad Mehdi Saber: Formal analysis, Investigation, Methodology, Project administration, Writing-review & editing.

Mohammad Reza Baneshi: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Writing-review & editing supervision.

- Acknowledgements

We would like to show our gratitude to Maryam Chegeni for sharing their pearls of wisdom with us during this research, and we thank “anonymous” reviewers for their insights.

Comparison of performance of recursive partitioning models and parametric imputation models based on predictive mean matching in multiple imputation by chained equations, in order to impute the missing values of the binary outcome, in the presence of an interaction effect

Abstract

Background: Among the new multiple imputation methods, Multiple Imputation by Chained Equations (MICE) is a popular approach for implementing multiple imputations because of its flexibility. Our main focus in this study is to compare the performance of parametric imputation models based on predictive mean matching and recursive partitioning methods in multiple imputation by chained equations in the presence of interaction in the data.

Methods: We compared the performance of parametric and tree-based imputation methods via simulation using two data generation models. For each combination of data generation model and imputation method, the following steps were performed: data generation, removal of observations, imputation, logistic regression analysis, and calculation of bias, Coverage Probability (CP), and Confidence Interval (CI) width for each coefficient. Furthermore, model-based and empirical SE, and estimated proportion of the variance attributable to the missing data (λ) were calculated.

Results: We have shown by simulation that to impute a binary response in observations involving an interaction, manually entering the interaction term into the imputation model in the predictive mean matching model improves the performance of the PMM method compared to the recursive partitioning models in multiple imputation by chained equations. The parametric method in which we entered the interaction model into the imputation model (MICE-Interaction) led to smaller bias, slightly higher coverage probability for the interaction effect, but it had slightly wider confidence intervals than tree-based imputation (especially classification and regression trees).

Conclusions: The application of MICE-Interaction led to better performance than recursive partitioning methods in MICE. Although the user is interested in estimating the interaction and does not know enough about the structure of the observations, recursive partitioning methods can be suggested to impute the missing values.

Keywords: Imputation, Interaction, Missing data, Regression tree, Predictive Mean Matching, Recursive Partitioning

Background

Among imputation methods, Multiple Imputation (MI) has received particular attention in different areas of research (Sterne et al., 2009, Rubin, 1987). The popularity of this method is attributed to its flexibility, besides its ability to produce unbiased estimates and valid inferences for data that are Missing At Random (MAR, i.e., when, conditional on the observed data, the probability of data being missing does not depend on unobserved data) and Missing Completely At Random (MCAR, i.e., when the probability of having a variable with missing data does not depend on any observed or missing variables) (Lee and Carlin, 2010).

Multiple Imputation by Chained Equations (MICE) is commonly used for imputing missing data (Raghunathan et al., 2001, Buuren and Groothuis-Oudshoorn, 2010). The primary advantage of MICE is that it is not necessary to specify a joint distribution for all variables (Gelfand and Smith, 1990). Any interactions or non-linear relationships that are estimated in the final analysis model must be accounted for in the imputation model. Incorrect selection of the imputation model in MICE may lead to invalid inferences. When there is an interaction between the variables, the default imputation methods (imputation models that include only the main effects) may produce biased parameter estimations since those imputation models are not compatible with the analysis model (Van Buuren, 2007, Seaman et al., 2012a, Slade and Naylor, 2020).

In some cases, it may not be desirable or possible to include interactions in a parametric imputation model due to numerous predictors, small sample size, or highly correlated predictors. In these situations, recursive partitioning (tree-based) methods, such as Classification And Regression Trees (CART) and Random Forests (RF), can be used within the MICE algorithm (Strobl et al., 2009).

CART and RF have some advantages due to their nonparametric inherent nature that is, they do not require the user to specify an imputation model, as well as these models, are flexible enough to capture the interaction effects and non-linear relations. At each step in the MICE algorithm, missing values are imputed for a given incomplete variable using a tree built with all other variables as predictors. A primary drawback of tree-based methods is the difficulty interpreting the results. However, this disadvantage is inconsequential for the imputation process (Strobl et al., 2009, Burgette and Reiter, 2010, Doove et al., 2014).

In this research, we aimed to conduct two simulation studies to compare the performance of (partially) parametric imputation methods (considering interaction in the imputation model) with recursive partitioning methods to find a more convenient method to preserve interaction effects.

This paper is organized as follows. At first, simulation designs are described. Then four imputation methods within the MICE algorithm were carried out to investigate, followed by the results of the simulation studies. The results from both simulation studies will be discussed in the last section.

Methods

We compared the performance of parametric and tree-based imputation methods via simulation using two data generation models. For each combination of data generation model and imputation method, the following steps were performed: data generation, removal of observations based on a MAR and MCAR mechanisms, imputation, logistic regression analysis, and calculation of bias, Coverage Probability (CP), and Confidence Interval (CI) width for each coefficient. In general, a desirable imputation method results in low Bias, CP of at least 95%, and narrow CIs. Furthermore, model-based and empirical SE, and estimated proportion of the variance attributable to the missing data (λ) were calculated.

Data generation

A total of 1000 simulated datasets with 1000 observations were generated, using two different logistic regression models, one with an interaction between two continuous variables and one with an interaction between two binary variables. The models are specified in Equations 1 and 2, respectively.

$$\log \text{it } [p(Y_1=1)] = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 X_1 + \beta_4 X_2 + \beta_5 X_3 + \beta_6 X_4 + \beta_7 X_5 + \beta_8 X_1 X_2 \quad (1)$$

$$\log \text{it } [p(Y_2=1)] = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + \gamma_3 X_1 + \gamma_4 X_2 + \gamma_5 X_3 + \gamma_6 X_4 + \gamma_7 X_5 + \gamma_8 Z_1 Z_2 \quad (2)$$

We simulated data from a multivariate normal distribution where all continuous variables had a mean of 0 and a standard deviation of 1 (X_1 to X_5 represent continuous variables) with the following correlation structure: $\text{Corr}(X_1, X_2) = \text{Corr}(X_1, X_3) = \text{Corr}(X_2, X_3) = 0.5$, $\text{Corr}(X_4, X_5) = 0.3$, $\text{Corr}(X_3, X_4) = 0$. Also, Z_1 and Z_2 are binary variables, that are independent of each other, randomly drawn from a binomial distribution. Two binary predictors are independent of the other predictors. The values of the model parameters were selected so that the intercepts were equal to zero, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = 0.25$, $\beta_3 = \dots = \beta_7 = 0.5$, $\gamma_3 = \dots = \gamma_7 = 0.5$, and $\beta_8 = \gamma_8 = 1$.

Removal of observations

In each of the datasets, we induced missing values at MAR and MCAR mechanisms such that the proportions of missing data of the outcome variable are 10%, 20%, 30%, 40%, and 50%.

Imputation of missing data

Throughout the analysis, we performed MICE using the mice package in R version 3.1.0 (Buuren and Groothuis-Oudshoorn, 2010, Team, 2013). Using PMM-based multiple multiple imputation methods and two recursive partitioning methods in MICE the missing values were imputed.

PMM

PMM is generally preferred over standard regression because it creates imputations from the observed data itself, which maintains data structure such as skewness and avoids problems such as imputing impossible values (Morris et al., 2014, Slade and Naylor, 2020). The default implementation of PMM in mice includes

only main effects in the imputation model. PMM corresponds to standard PMM presented in Doove et al. (Doove et al., 2014).

MICE-Interaction

For MICE-Interaction, we included the interaction term as just another variable and used the default predictor matrix in the mice function in R software (Von Hippel, 2009, Seaman et al., 2012b). The difference between MICE-Interaction and PMM (MICE-PMM) is the inclusion of an interaction term in the imputation model for MICE-Interaction.

MICE-Stratified

For the MICE-Stratified method, the data sets are stratified by the fully observed binary variables that contributed to the interactions then the PMM model in the MICE method is used to impute missing values in two subgroups.

MICE-CART

CART is a tree-based imputation method that does not require the specification of an imputation model. In short, the decision tree is created by CART, based on a binary decision rule, using one predictor variable that partitions the data into two nodes by minimizing the variance of the outcome within each node. Subgroups are created by the optimal split, according to a measure of homogeneity, such as the Gini index (Breiman L, 1984). To avoid overfitting, partitioning continues until a certain criterion is obtained (e.g., a predetermined number of observations in the final subsets) (Friedman et al., 2001, Breiman L, 1984).

The data matrix Y is partitioned into Y^{obs} and Y^{miss} . Y^{obs} includes the columns of Y that have been completely observed, and Y^{miss} is part of Y that has not been completely observed. The following steps describe the MICE-CART procedure (k is the number of incomplete variables, and $\hat{Y}$ is the currently imputed data matrix Y) (Burgette and Reiter, 2010):

1. For $j = 1, \dots, k$, the initial values of Y_j are filled by random draws from Y_j^{obs} , and a data matrix Z is defined.
2. The CART is fitted on each Y_j^{miss} by using the remaining variables of Z as the predictor variables; only subjects with observed values on Y_j^{miss} are used in this process.
3. For subjects in Y_j^{miss} , the terminal node is determined, according to the fitted tree in step 2. An observed value on Y_j^{miss} is randomly selected from the subset in this node and used for imputation.
4. Step 2 and step 3 are repeated for several iterations. This procedure is performed for each variable with at least one missing value, yielding one complete dataset.
5. Steps 1-4 are repeated several times (L), resulting in L imputed datasets (Doove et al., 2014).

For the CART method in MICE, the complexity parameter of 10^{-4} and a minimum of five observations at any terminal node were applied (Therneau et al., 2018, Buuren and Groothuis-Oudshoorn, 2010).

MICE-RF

The RF algorithm is essentially an extension of CART that creates multiple trees, using bootstrap samples from the original data. To overcome overfitting, at each node, a random sample of variables is used to determine the optimal split. The RF imputation method creates multiple regression trees by randomly selecting B bootstrap samples from the original sample (Hayes et al., 2015). To make each split, a subsample of independent variables is selected randomly and used with CART to produce a tree (Breiman, 2001). If we consider the same description for CART in the MICE section, instead of generating a tree over the entire observations, B bootstrap samples are first extracted; then, a tree is fitted on each bootstrap sample. Each tree contains several leaves, and each leaf contains a subset of observed values for variable j , called donors. Then, the leaf associated with the missing value in the j^{th} variable (Y_j^{miss}) is checked, based on the B node of the created tree (Breiman, 2001).

For each Y_j^{miss} , if we consider all donors in the B node of the resulting leaf, a random value is selected from the donors. For each variable with missing values, this process is repeated to produce a complete observation set. The rest of the imputation process is the same as the CART method in MICE (Doove et al., 2014). In the present study, we used the mice package to implement RF in MICE (Buuren and Groothuis-Oudshoorn, 2010, Liaw and Wiener, 2002). We drew 10 bootstrap samples (White et al., 2011, Shah et al., 2014) to make each split candidate one-third of the predictors (Buuren and Groothuis-Oudshoorn, 2010, Doove et al., 2014). However, we did not consider changing the number of trees (n_{tree}), because the literature (Shah et al., 2014) suggested that the imputation quality is equivalent for $n_{\text{tree}} = 10$ and $n_{\text{tree}} = 100$.

Regression analysis

For each of the 1000 imputed datasets, we fit a correctly specified final analysis model, logistic models, and the results are combined using Rubin's rules (Rubin, 2004).

Calculation of bias, CP, and CI width

We computed Bias as the estimated coefficient minus the true value. For a single repetition of the simulation, CP (Mistler and Enders) is 0 if the estimated 95% CI does not contain the true value and 1 if the estimated interval does contain the truth. In 1000 iterations of the simulation, the CP is the proportion of the confidence intervals, which contains the true value of the desired coefficient. Average CI Length (AL) refers to the width of the estimated 95% CI for each coefficient. We report the discussed measurements across the 1000 replications for each of the imputation methods. Additionally, we computed empirical standard error for each coefficient, which is computed as the SD of the estimates across 1000 simulation replications. (Doove et al., 2014) (Shah et al., 2014).

Results

Scenario 1 (Interaction between two continuous variables)

For this scenario, Table 1 displays the mean values of biases across the 1000 replications, and Tables 2 and 3 display the coverage probabilities and average 95% confidence interval widths, respectively, for each imputation method. For the interaction effect, MICE-Stratified led to estimates with higher mean biases than the tree-based methods. Further, the CP values of the interaction effects at all missing percentages are the smallest using MICE-Stratified (For instance, under the MAR mechanism, it was only 33.8% at 10% missing percentage).

For the interaction effect, MICE-Interaction had smaller mean absolute biases and larger coverage probability values at all the missing percentages (>99.0%) than the tree-based methods. The mean absolute values of relative biases in estimating the interaction at 20% and 30% missing percentages under the MAR mechanism were, respectively, 0.001 and 0.006 for the MICE-Interaction method, and it generated complete coverage values.

For the MICE-Interaction method, the mean absolute values of relative biases for the coefficients contributing to interaction (X_1 and X_2), at 20% missing percentage under the MAR mechanism were 0.010 and 0.006, respectively, and were 0.002 and 0.005 at a 30% missing percentage. Furthermore, the coverage values for these coefficients, at 20% and 30% missing percentages under the MAR mechanism, were complete or very close to one.

The mean absolute values of the relative biases in estimating all coefficients for the MICE-CART method were less than MICE-RF and MICE-Stratified (for instance, the mean bias of the variable X_2 at 20% missing percentage under MAR, for MICE-CART method was -0.041, for MICE-RF method is -0.103 and for MICE-Stratified was equal to 0.200). In 80% of cases, the MICE-CART method led to mean absolute values of the relative biases less than 0.20 and resulted in acceptable coverage values. In almost all cases, the standard errors, and variations attributed to the missing values of the MICE-CART model were less than the MICE-RF model.

Considering the results of relative biases, coverages, and λ values, the MICE-Interaction method was recognized as superior. The model standard error values and the empirical standard error were close to each other, so the variances estimated by the Rubin rules were unbiased.

Across replications of this simulation, at 10% and 20% missing percentages, the absolute biases of the main effects from the MICE-RF and MICE-CART methods had narrower widths than those from the MICE-Interaction method. Thus, despite the fact that the MICE-RF and MICE-CART methods led to point estimates with higher mean bias than MICE-Interaction, they maintained valid CP for the main effects with the narrowest CIs. We observed this phenomenon for all main effects in Scenario 1, but as shown in Tables 1-3, this phenomenon did not occur for the interaction effect.

CP values for the MICE-Interaction method were greater than 0.95. For 40% and 50% missing percentages, the CP values for the recursive partitioning methods were less than 0.95 and were not acceptable at all. The MICE-Stratified method produced less than 0.95 for almost all effects of CP values.

In the presence of interaction between two continuous variables, for the main effects, MICE-CART led to smaller SE than MICE-Interaction and MICE-RF.

Scenario 2 (Interaction between two binary variables)

In the case where there is an interaction between two binary variables, Table 4 displays the mean values of biases across the 1000 replications, and Tables 5 and 6 display the coverage probabilities and average 95% confidence interval widths, respectively, for each imputation method. In general, the bias values for the MICE-Interaction method were lower than other methods. Recursive partitioning methods underestimated all coefficients except for one of the main effects. For such interaction, the recursive partitioning methods created bias values greater than 0.05.

For the MICE-Interaction method, the mean absolute values of the relative biases in estimating the interaction, at the missing percentages of 20% and 30% under the MAR mechanism, were -0.007 and -0.001, respectively. Furthermore, coverage values were 1 and 0.999, respectively. In estimating the interaction, the coverage values for all four methods were greater than 0.95 at 20% missing percentage under MAR, while at 30% missing percentage, the CP values of the MICE-Stratified method were lower than 0.95. At 40% and 50% missing percentages, the bias values of the recursive partitioning methods were greater than 0.40. In estimating the main effects, the MICE-Interaction method resulted in the least biases among the four methods.

For the MICE-Interaction method, the mean absolute values of the relative biases in estimating the coefficients contributed to the interaction term, Z_1 and Z_2 , at 20% missing percentage under the MAR mechanism, were -0.018 and -0.17 and at 30% missing percentage were 0.002- and 0.001, respectively.

For the MICE-CART method, the mean absolute values of the relative biases in estimating the coefficients non-contributing to the interaction were lower than the MICE-RF method. For instance, at a 30% missing percentage under MAR, in estimating the variable X_4 , the bias of the MICE-CART method was -0.65 vs. -0.125 for the MICE-RF method.

In estimating the main effects, the coverage values for all four methods at 20% and 30% missing percentages at the MAR mechanism were more than 0.95.

In 70% of cases, the MICE-CART method produced mean absolute values of the relative biases less than 0.20 and produced acceptable coverage values (almost 0.95 or at least 0.95). In almost all cases, the standard errors and variations attributed to the missing values of the MICE-CART model were less than the MICE-RF model.

Although the length of the CI, the standard error of the model, and the ratio of variations attributed to the MICE-Stratified method were the lowest, but due to the values of bias, relative bias and, CP, this method was not the acceptable model.

The values of the empirical standard error and the standard error of the model were almost the same, indicating that the variance of the coefficients obtained from the Rubin rules was unbiased.

Discussion

The CART and RF models in MICE were evaluated in comparison with the compatible parametric model (MICE-Interaction) to investigate whether recursive partitioning methods can preserve interactions and main effects. Our simulated datasets included a binary outcome and a combination of continuous and binary predictors. At all missing percentages, the compatible parametric model (MICE-Interaction) led to estimating the true interaction effect with lower absolute relative biases and higher CPs than the tree-based methods. Omitting the true interaction term from the parametric imputation model resulted in estimating the interaction effect with the largest mean bias and smallest CP values (Slade and Naylor, 2020).

Recent research has shown that whenever the imputation model and the analysis model were compatible, the parametric imputation model was superior to nonparametric tree-based imputation methods (CART and RF) (Slade and Naylor, 2020). However, no study has compared the performance of these methods in binary responses, when both binary and continuous predictors are available.

In the study conducted by Javadi et al. For the case where the target was imputing the response variable in the presence of an interaction between a binary and continuous predictors, the four discussed imputation methods were compared, and they concluded that by manual entry interaction term in the PMM model, the performance of PMM-based imputation methods compared to recursive partitioning methods, considering bias, coverage, and variation attributed by missing values were highly improved (Javadi et al., 2021).

In all scenarios, we calculated the values of the variation attributed to the missing values (λ values). Due to the fact that all the computed λ values were almost equal to 0.50, we refused to report them. λ values less than half indicate that the imputation results were less affected by the imputation model used and more affected by the observed values (Van Buuren, 2018, Doove et al., 2014).

In this study, in all scenarios with an interaction between two binary variables, MICE-Interaction was superior to other methods. This means that in addition to the interaction estimates, MICE-Interaction estimated all the main effects with the least amount of bias values, and acceptable coverages.

In scenarios with an interaction between two quantitative variables, in about 80% of cases, the MICE-CART method resulted in the mean absolute of the relative biases less than 0.20. In estimating the interaction, the coverage values of the MICE-Interaction were acceptable in all scenarios. In this case, at the missing percentages of more than 30%, the MICE-CART method should be used with caution since the coverage values were less than 0.95.

In the case where there was an interaction between two binary variables, although the performance of the MICE-Interaction method was much better than other methods, but in estimating the interaction, tree-based methods also had acceptable performance. In about 70% of cases, the MICE-CART method resulted in absolute relative bias values less than 0.20, and in about 90% of cases, the coverage probability values were close to 0.95 or at least 0.95.

The present study has several strengths. For the first time, we compared tree-based imputation methods (CART and RF) with parametric imputation methods in MICE by specifying a parametric imputation model (MICE-Interaction) in simulation studies with binary outcomes, in the presence of interactions. Also, we used the simulation technique to generate missing data in complete datasets; therefore, the true value of the missing data was known for evaluating multiple imputation methods; this is important because we had access to the original observations before creating the missing values and could compare the results of real and imputed observations. Moreover, the missing percentages were generated under two mechanisms on binary response variables.

Some limitations of this study should be considered. Firstly, only one condition was considered for the MAR mechanism, which was also arbitrary; therefore, the results can only be generalized to this condition, not the rest of the MAR mechanisms. Secondly, it would be preferable if we had access to true data and could evaluate the performance of methods on these data. Finally, we considered the sample size to be 1000.

Overall, the selection of predictors and their effects for incorporation in the final model is an issue beyond the scope of this study. Our recommendations for the imputation model are based on the assumption that the final model is correctly defined. Future research should consider the interplay between the misspecification of the final model and the selection of MICE models.

The number of trees built for the RF model was set to 10 to limit bias and compared to PMM operating under their default parameters. The iteration number for imputation methods based on MICE was set to five as suggested. But in practice, the influence of tuning parameters on imputation performance and computational burden still and further study is warranted, and parameter searching may be necessary for complex datasets to attain better imputation performance. For CART, the default value of 10^{-4} is appropriate. Smaller values of minbucket lead to slightly decreased mean bias, but the default value of 5 seemed reasonable (Breiman, 2001, Doove et al., 2014) (Slade and Naylor, 2020).

Conclusion

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 and a binary outcome. If interest lies primarily in the estimation of an Interaction between two continuous variables, we recommend utilizing MICE-Interaction as it leads to the lowest bias, the highest CP, and narrow 95% CIs for the interaction effects at all missing percentages. In this case, If interest lies primarily in the estimation of main effects, there is a trade-off between tree-based and parametric imputations at missing percentages below 30%. MICE-Interaction had the highest CP of the interaction effect, but it was at the expense of wide 95% CIs. In the presence of interaction between two binary variables, we recommend using MICE-Interaction as it led to the lowest bias, acceptable CP for the interaction effects at all missing percentages. Importantly, parametric imputation should only be utilized if there is enough information to ensure that all necessary interaction terms were included in the imputation model.

Our results have important implications for epidemiologists in dealing with incomplete data. Researchers should consider that in order to reduce the biases of the estimated coefficients if they are aware of the interactions, those interactions should be included in the parametric PMM imputation model to make

convergence between the imputation model and the analysis model. Therefore, if the researcher has sufficient information about the structure of the data and the user is able to correctly detect and identify nonlinear effects such as interaction, the use of the MICE-Interaction method is definitely recommended. Of course, in such a situation, the results obtained from the MICE-Interaction method are ideal results and may not actually occur because it is challenging and difficult to accurately identify all the interactions and relationships in the observations. Therefore, if the researcher is interested in estimating the interaction and does not have enough information about the structure of the observations, it is better to evaluate the recursive partitioning methods in MICE because these models detect interaction and nonlinear effects automatically and inherently. It is not recommended to use the PMM method in MICE, which is the default of most statistical software.

List of abbreviations

MCR: MICE-CART; MRF: MICE-RF; MI: MICE-Interaction; MS: MICE -Stratified; MAR: Missing At Random; MCAR: Missing Completely At Random; MI: Multiple Imputation; PMM: Predictive Mean Matching; RF: Random Forests; SE: Standard Error; SD: Standard Deviation

Table 1 Scenario 1: Bias of each coefficient.

Variable	Model	MAR					MCAR				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
X_1	MCR	-0.022	-0.031	-0.056	-0.067	-0.093	-0.015	-0.031	-0.045	-0.066	-0.079
	MI	-0.001	0.005	0.001	0.005	0.002	0.001	0.000	0.001	0.002	0.003
	MRF	-0.054	-0.096	-0.142	-0.175	-0.211	-0.052	-0.099	-0.138	-0.178	-0.209
	MS	-0.116	-0.192	-0.255	-0.304	-0.346	-0.118	-0.204	-0.267	-0.317	-0.357
$X_1 * X_2$	MCR	-0.057	-0.112	-0.168	-0.228	-0.281	-0.053	-0.110	-0.169	-0.230	-0.289
	MI	0.000	0.001	0.006	0.006	0.006	0.003	0.006	0.002	0.009	0.008
	MRF	-0.119	-0.229	-0.323	-0.413	-0.494	-0.121	-0.226	-0.324	-0.413	-0.496
	MS	-0.231	-0.399	-0.522	-0.620	-0.703	-0.235	-0.405	-0.536	-0.638	-0.725
X_2	MCR	-0.018	-0.041	-0.057	-0.072	-0.093	-0.015	-0.032	-0.051	-0.065	-0.084
	MI	0.001	-0.003	0.002	0.002	0.000	0.000	0.004	-0.002	0.000	0.002
	MRF	-0.051	-0.103	-0.140	-0.177	-0.213	-0.054	-0.099	-0.143	-0.179	-0.213
	MS	-0.115	-0.200	-0.258	-0.306	-0.348	-0.120	-0.206	-0.269	-0.322	-0.361
X_3	MCR	-0.014	-0.027	-0.036	-0.061	-0.068	-0.014	-0.027	-0.042	-0.054	-0.072
	MI	0.002	0.000	0.007	0.004	0.010	-0.001	0.002	0.002	0.001	0.004
	MRF	-0.040	-0.077	-0.105	-0.139	-0.168	-0.041	-0.074	-0.107	-0.136	-0.164
	MS	-0.087	-0.156	-0.208	-0.259	-0.305	-0.086	-0.151	-0.209	-0.262	-0.307
X_4	MCR	-0.021	-0.036	-0.052	-0.068	-0.090	-0.011	-0.024	-0.036	-0.050	-0.061
	MI	-0.002	-0.001	0.001	0.002	-0.001	0.001	0.002	0.002	0.001	0.006
	MRF	-0.047	-0.086	-0.119	-0.154	-0.190	-0.038	-0.074	-0.107	-0.138	-0.166
	MS	-0.111	-0.182	-0.229	-0.272	-0.307	-0.082	-0.151	-0.208	-0.260	-0.304
X_5	MCR	-0.020	-0.038	-0.058	-0.075	-0.091	-0.011	-0.022	-0.036	-0.052	-0.060
	MI	-0.002	-0.003	-0.003	-0.003	-0.006	0.002	0.003	0.002	0.002	0.005
	MRF	-0.046	-0.086	-0.125	-0.159	-0.192	-0.038	-0.073	-0.106	-0.138	-0.163
	MS	-0.113	-0.180	-0.234	-0.274	-0.310	-0.082	-0.147	-0.208	-0.261	-0.305
Z_1	MCR	-0.022	-0.040	-0.065	-0.079	-0.103	-0.023	-0.046	-0.065	-0.090	-0.109
	MI	-0.003	-0.003	-0.004	-0.005	0.002	0.001	0.000	-0.001	-0.002	-0.007
	MRF	-0.026	-0.048	-0.071	-0.088	-0.108	-0.024	-0.046	-0.065	-0.090	-0.111
	MS	-0.043	-0.073	-0.105	-0.130	-0.151	-0.042	-0.077	-0.102	-0.129	-0.153
Z_2	MCR	-0.021	-0.040	-0.064	-0.073	-0.108	-0.020	-0.044	-0.058	-0.087	-0.098
	MI	-0.002	0.002	-0.006	0.005	-0.003	-0.001	0.000	0.006	-0.002	0.004
	MRF	-0.025	-0.045	-0.074	-0.086	-0.113	-0.026	-0.045	-0.066	-0.090	-0.103
	MS	-0.048	-0.077	-0.111	-0.136	-0.157	-0.044	-0.085	-0.107	-0.137	-0.158

Table 2 Scenario 1: CP of each coefficient.

Variable	Model	MAR					MCAR				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
X_1	MCR	1.000	0.999	0.972	0.928	0.844	1.000	0.994	0.972	0.922	0.843
	MI	1.000	1.000	1.000	0.996	0.980	1.000	1.000	0.998	0.993	0.983
	MRF	1.000	0.988	0.920	0.787	0.649	1.000	0.980	0.893	0.779	0.666
	MS	0.908	0.434	0.141	0.052	0.019	0.888	0.393	0.121	0.037	0.016
$X_1 * X_2$	MCR	0.997	0.949	0.766	0.560	0.391	0.998	0.951	0.780	0.550	0.409
	MI	1.000	1.000	1.000	0.997	0.991	1.000	1.000	1.000	0.997	0.989
	MRF	0.987	0.619	0.214	0.063	0.014	0.997	0.668	0.236	0.067	0.018
	MS	0.338	0.006	0.000	0.000	0.000	0.311	0.004	0.000	0.000	0.000
X_2	MCR	1.000	0.996	0.973	0.908	0.873	1.000	0.994	0.970	0.928	0.858
	MI	1.000	1.000	0.997	0.996	0.992	1.000	1.000	0.999	0.996	0.988
	MRF	1.000	0.982	0.914	0.778	0.659	1.000	0.988	0.901	0.814	0.644
	MS	0.909	0.400	0.153	0.043	0.016	0.891	0.377	0.106	0.038	0.011
X_3	MCR	1.000	0.997	0.972	0.936	0.886	1.000	0.999	0.983	0.927	0.858
	MI	1.000	1.000	0.998	0.994	0.987	1.000	1.000	0.997	0.996	0.996
	MRF	1.000	0.996	0.966	0.882	0.787	1.000	0.996	0.961	0.907	0.789
	MS	0.979	0.640	0.297	0.118	0.044	0.979	0.671	0.294	0.111	0.033
X_4	MCR	1.000	0.987	0.956	0.868	0.770	1.000	0.990	0.979	0.927	0.894
	MI	1.000	0.999	0.999	0.987	0.988	1.000	1.000	0.999	0.997	0.983
	MRF	1.000	0.971	0.911	0.723	0.558	1.000	0.987	0.920	0.821	0.707
	MS	0.839	0.273	0.092	0.032	0.012	0.967	0.523	0.178	0.053	0.012
X_5	MCR	1.000	0.988	0.931	0.854	0.781	1.000	0.994	0.977	0.940	0.879
	MI	1.000	1.000	0.997	0.990	0.989	1.000	1.000	0.998	0.996	0.982
	MRF	1.000	0.977	0.846	0.701	0.558	1.000	0.992	0.939	0.802	0.703
	MS	0.826	0.292	0.071	0.024	0.008	0.962	0.533	0.151	0.047	0.006
Z_1	MCR	1.000	1.000	0.993	0.984	0.961	1.000	0.999	0.991	0.980	0.962
	MI	1.000	1.000	0.999	0.996	0.986	1.000	1.000	1.000	0.988	0.984
	MRF	1.000	0.999	0.999	0.993	0.990	1.000	1.000	0.998	0.991	0.986
	MS	1.000	0.992	0.954	0.930	0.909	0.999	0.989	0.956	0.926	0.899
Z_2	MCR	1.000	1.000	0.997	0.981	0.968	1.000	0.999	0.997	0.986	0.969
	MI	1.000	1.000	0.998	0.993	0.990	1.000	1.000	1.000	0.994	0.987
	MRF	1.000	1.000	0.999	0.994	0.989	1.000	1.000	0.998	0.996	0.990
	MS	1.000	0.989	0.954	0.914	0.890	1.000	0.976	0.952	0.931	0.891

Table 3 Scenario1: 95% CI width of each coefficient.

Variable	Model	MAR					MCAR				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
X_1	MCR	0.421	0.437	0.447	0.461	0.473	0.424	0.440	0.455	0.467	0.478
	MI	0.435	0.467	0.515	0.566	0.644	0.438	0.476	0.520	0.587	0.670
	MRF	0.433	0.452	0.472	0.481	0.496	0.435	0.459	0.476	0.487	0.502
	MS	0.381	0.367	0.362	0.363	0.370	0.383	0.371	0.366	0.367	0.371
$X_1 * X_2$	MCR	0.460	0.481	0.490	0.500	0.499	0.465	0.482	0.498	0.504	0.514
	MI	0.467	0.501	0.547	0.610	0.674	0.472	0.513	0.566	0.627	0.738
	MRF	0.484	0.514	0.517	0.516	0.510	0.489	0.516	0.526	0.525	0.525
	MS	0.387	0.358	0.341	0.329	0.320	0.391	0.364	0.345	0.333	0.326
X_2	MCR	0.422	0.436	0.450	0.464	0.476	0.425	0.438	0.456	0.464	0.479
	MI	0.434	0.467	0.512	0.578	0.647	0.438	0.476	0.528	0.584	0.673
	MRF	0.432	0.454	0.471	0.488	0.497	0.436	0.459	0.472	0.490	0.502
	MS	0.381	0.367	0.362	0.362	0.369	0.383	0.370	0.365	0.367	0.372
X_3	MCR	0.392	0.405	0.418	0.428	0.444	0.393	0.408	0.417	0.433	0.442
	MI	0.403	0.435	0.476	0.532	0.619	0.405	0.440	0.487	0.546	0.613
	MRF	0.399	0.417	0.437	0.456	0.478	0.401	0.420	0.441	0.457	0.471
	MS	0.361	0.352	0.351	0.354	0.363	0.363	0.357	0.356	0.358	0.365
X_4	MCR	0.343	0.354	0.365	0.376	0.385	0.341	0.353	0.365	0.374	0.385
	MI	0.355	0.386	0.431	0.478	0.549	0.351	0.382	0.419	0.468	0.532
	MRF	0.352	0.371	0.387	0.395	0.412	0.347	0.364	0.379	0.397	0.416
	MS	0.311	0.304	0.304	0.306	0.312	0.314	0.308	0.305	0.310	0.315
X_5	MCR	0.342	0.355	0.364	0.375	0.384	0.341	0.354	0.365	0.374	0.386
	MI	0.354	0.387	0.432	0.484	0.553	0.352	0.381	0.422	0.469	0.533
	MRF	0.353	0.370	0.383	0.397	0.409	0.347	0.366	0.381	0.395	0.413
	MS	0.311	0.304	0.303	0.306	0.314	0.314	0.307	0.305	0.309	0.314
Z_1	MCR	0.618	0.639	0.658	0.675	0.691	0.621	0.642	0.661	0.679	0.706
	MI	0.636	0.689	0.760	0.850	0.967	0.642	0.698	0.768	0.860	0.992
	MRF	0.629	0.659	0.687	0.716	0.750	0.632	0.664	0.702	0.725	0.764
	MS	0.576	0.566	0.565	0.573	0.587	0.580	0.571	0.572	0.581	0.595
Z_2	MCR	0.620	0.637	0.657	0.675	0.694	0.621	0.642	0.661	0.681	0.706
	MI	0.640	0.689	0.756	0.848	0.948	0.643	0.702	0.774	0.851	0.977
	MRF	0.630	0.657	0.687	0.715	0.752	0.635	0.664	0.700	0.730	0.758
	MS	0.576	0.566	0.566	0.574	0.588	0.580	0.572	0.572	0.580	0.596

Table 4 Scenario 2: Bias of each coefficient.

Variable	Model	MAR					MCAR				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
X ₁	MCR	-0.010	-0.023	-0.029	-0.034	-0.047	-0.010	-0.017	-0.030	-0.040	-0.051
	MI	0.002	0.001	0.005	0.008	0.012	0.002	0.001	0.001	0.005	0.005
	MRF	-0.035	-0.071	-0.100	-0.128	-0.155	-0.035	-0.069	-0.099	-0.126	-0.154
	MS	-0.084	-0.156	-0.211	-0.258	-0.304	-0.082	-0.154	-0.214	-0.260	-0.310
X ₂	MCR	-0.011	-0.020	-0.030	-0.042	-0.053	-0.010	-0.019	-0.024	-0.039	-0.042
	MI	0.001	0.003	0.005	0.006	0.005	0.000	0.003	0.009	0.005	0.015
	MRF	-0.038	-0.072	-0.102	-0.131	-0.162	-0.035	-0.067	-0.094	-0.127	-0.149
	MS	-0.086	-0.152	-0.209	-0.259	-0.306	-0.086	-0.150	-0.210	-0.265	-0.311
X ₃	MCR	-0.011	-0.020	-0.029	-0.036	-0.056	-0.010	-0.020	-0.030	-0.040	-0.055
	MI	-0.001	0.000	0.003	0.013	0.014	0.001	0.001	0.001	0.008	0.009
	MRF	-0.038	-0.071	-0.101	-0.128	-0.159	-0.036	-0.069	-0.103	-0.129	-0.158
	MS	-0.089	-0.155	-0.212	-0.263	-0.308	-0.087	-0.156	-0.214	-0.266	-0.310
X ₄	MCR	-0.023	-0.049	-0.065	-0.084	-0.106	-0.015	-0.029	-0.040	-0.056	-0.075
	MI	0.000	-0.002	0.003	0.004	0.009	-0.001	0.001	0.004	0.007	0.009
	MRF	-0.047	-0.092	-0.125	-0.162	-0.198	-0.041	-0.075	-0.109	-0.139	-0.170
	MS	-0.110	-0.184	-0.233	-0.273	-0.311	-0.086	-0.154	-0.215	-0.263	-0.312
X ₅	MCR	-0.025	-0.048	-0.065	-0.088	-0.108	-0.012	-0.026	-0.045	-0.060	-0.069
	MI	0.001	-0.001	0.001	-0.001	0.001	0.002	0.003	0.004	0.006	0.017
	MRF	-0.047	-0.091	-0.126	-0.163	-0.199	-0.037	-0.075	-0.109	-0.143	-0.168
	MS	-0.110	-0.181	-0.233	-0.274	-0.308	-0.082	-0.151	-0.211	-0.263	-0.307
Z ₁	MCR	0.077	0.172	0.256	0.313	0.439	0.104	0.208	0.259	0.369	0.452
	MI	0.001	0.014	0.002	-0.015	-0.006	0.003	0.015	-0.028	-0.013	0.000
	MRF	0.090	0.188	0.273	0.351	0.432	0.100	0.198	0.265	0.348	0.442
	MS	0.136	0.248	0.315	0.390	0.464	0.125	0.252	0.321	0.413	0.477
Z ₁ *Z ₂	MCR	-0.087	-0.184	-0.282	-0.355	-0.470	-0.109	-0.216	-0.288	-0.402	-0.491
	MI	0.000	-0.007	-0.001	0.015	0.014	-0.002	-0.008	0.020	0.012	0.010
	MRF	-0.107	-0.218	-0.319	-0.407	-0.498	-0.116	-0.224	-0.311	-0.406	-0.505
	MS	-0.178	-0.324	-0.427	-0.524	-0.616	-0.169	-0.324	-0.431	-0.542	-0.628
Z ₂	MCR	0.080	0.169	0.257	0.311	0.430	0.105	0.209	0.263	0.371	0.459
	MI	0.005	0.012	0.001	-0.023	-0.010	0.005	0.014	-0.027	-0.011	-0.006
	MRF	0.095	0.186	0.275	0.347	0.431	0.099	0.196	0.266	0.353	0.440
	MS	0.144	0.255	0.318	0.391	0.465	0.122	0.254	0.326	0.414	0.476

Table 5 Scenario 2: CP of each coefficient.

Variable	Model	MAR					MCAR				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
X_1	MCR	1.000	0.993	0.981	0.946	0.907	1.000	0.995	0.989	0.956	0.904
	MI	1.000	1.000	0.999	0.997	0.990	1.000	1.000	0.999	0.994	0.981
	MRF	1.000	0.999	0.985	0.945	0.894	1.000	0.998	0.990	0.957	0.904
	MS	0.986	0.788	0.489	0.304	0.159	0.991	0.788	0.493	0.310	0.152
X_2	MCR	1.000	1.000	0.987	0.952	0.900	1.000	0.998	0.987	0.950	0.913
	MI	1.000	1.000	0.998	0.999	0.989	1.000	0.999	0.997	0.988	0.989
	MRF	1.000	0.999	0.982	0.953	0.900	1.000	0.998	0.982	0.943	0.907
	MS	0.987	0.804	0.505	0.301	0.143	0.990	0.814	0.527	0.273	0.159
X_3	MCR	1.000	1.000	0.983	0.949	0.896	1.000	0.996	0.979	0.956	0.878
	MI	1.000	1.000	0.998	0.994	0.986	1.000	1.000	1.000	0.994	0.987
	MRF	1.000	0.998	0.979	0.954	0.895	1.000	0.998	0.986	0.949	0.902
	MS	0.983	0.793	0.501	0.270	0.138	0.982	0.775	0.488	0.262	0.155
X_4	MCR	1.000	0.988	0.945	0.883	0.784	1.000	0.997	0.986	0.945	0.872
	MI	1.000	1.000	0.998	0.990	0.988	1.000	1.000	0.999	0.995	0.986
	MRF	1.000	0.987	0.933	0.819	0.697	1.000	0.994	0.963	0.913	0.795
	MS	0.920	0.507	0.230	0.123	0.061	0.975	0.678	0.319	0.140	0.058
X_5	MCR	0.999	0.985	0.936	0.884	0.788	1.000	0.998	0.977	0.942	0.876
	MI	1.000	1.000	0.997	0.994	0.987	1.000	1.000	0.999	0.998	0.986
	MRF	1.000	0.985	0.923	0.821	0.683	1.000	0.997	0.962	0.895	0.821
	MS	0.915	0.527	0.235	0.114	0.062	0.986	0.705	0.329	0.153	0.065
Z_1	MCR	1.000	1.000	0.996	0.990	0.972	1.000	1.000	0.995	0.978	0.969
	MI	1.000	1.000	0.999	0.998	0.987	1.000	1.000	0.999	0.995	0.987
	MRF	1.000	1.000	1.000	0.997	0.991	1.000	1.000	0.998	0.996	0.994
	MS	1.000	0.992	0.982	0.944	0.930	1.000	0.993	0.982	0.952	0.932
Z_1*Z_2	MCR	1.000	0.999	0.985	0.964	0.907	1.000	0.999	0.981	0.921	0.881
	MI	1.000	1.000	0.999	0.997	0.989	1.000	0.999	0.998	0.997	0.986
	MRF	1.000	1.000	0.995	0.974	0.944	1.000	1.000	0.996	0.975	0.929
	MS	0.999	0.961	0.859	0.736	0.613	1.000	0.952	0.851	0.710	0.614
Z_2	MCR	1.000	1.000	0.994	0.994	0.971	1.000	0.999	0.994	0.980	0.964
	MI	1.000	1.000	1.000	0.996	0.992	1.000	1.000	0.999	0.996	0.986
	MRF	1.000	1.000	1.000	0.995	0.988	1.000	1.000	0.997	0.995	0.989
	MS	1.000	0.994	0.982	0.955	0.928	1.000	0.993	0.979	0.948	0.927

Table 6 Scenario2: 95% CI width of each coefficient.

Variable	Model	MAR					MCAR				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
X ₁	MCR	0.475	0.491	0.509	0.525	0.547	0.477	0.494	0.511	0.530	0.546
	MI	0.490	0.530	0.578	0.652	0.733	0.492	0.537	0.591	0.659	0.750
	MRF	0.486	0.512	0.530	0.553	0.582	0.487	0.516	0.531	0.561	0.584
	MS	0.437	0.426	0.423	0.425	0.433	0.440	0.430	0.429	0.430	0.437
X ₂	MCR	0.474	0.492	0.507	0.523	0.542	0.476	0.495	0.511	0.528	0.545
	MI	0.490	0.527	0.578	0.647	0.747	0.492	0.535	0.588	0.660	0.760
	MRF	0.484	0.510	0.534	0.554	0.580	0.488	0.515	0.537	0.559	0.584
	MS	0.437	0.426	0.423	0.425	0.433	0.440	0.430	0.428	0.430	0.435
X ₃	MCR	0.475	0.492	0.508	0.526	0.535	0.478	0.494	0.511	0.526	0.548
	MI	0.490	0.532	0.579	0.650	0.739	0.492	0.533	0.597	0.661	0.751
	MRF	0.486	0.509	0.534	0.555	0.579	0.487	0.515	0.539	0.561	0.588
	MS	0.437	0.426	0.423	0.425	0.432	0.440	0.431	0.427	0.431	0.437
X ₄	MCR	0.415	0.429	0.441	0.456	0.463	0.412	0.427	0.444	0.454	0.469
	MI	0.428	0.466	0.518	0.582	0.669	0.427	0.465	0.504	0.576	0.667
	MRF	0.427	0.448	0.466	0.487	0.501	0.422	0.442	0.464	0.487	0.500
	MS	0.377	0.367	0.365	0.367	0.373	0.380	0.371	0.367	0.370	0.373
X ₅	MCR	0.414	0.430	0.443	0.454	0.464	0.414	0.428	0.443	0.460	0.469
	MI	0.430	0.470	0.522	0.586	0.671	0.427	0.464	0.511	0.568	0.654
	MRF	0.427	0.448	0.465	0.488	0.495	0.421	0.444	0.467	0.488	0.506
	MS	0.376	0.367	0.364	0.368	0.375	0.380	0.370	0.367	0.370	0.377
Z ₁	MCR	2.346	2.425	2.503	2.584	2.657	2.351	2.440	2.514	2.596	2.692
	MI	2.425	2.612	2.893	3.211	3.671	2.430	2.644	2.900	3.232	3.690
	MRF	2.388	2.502	2.621	2.722	2.805	2.398	2.539	2.631	2.768	2.854
	MS	2.182	2.140	2.137	2.157	2.196	2.195	2.159	2.157	2.176	2.212
Z ₁ *Z ₂	MCR	1.517	1.568	1.611	1.663	1.702	1.522	1.575	1.620	1.669	1.728
	MI	1.567	1.688	1.858	2.061	2.344	1.575	1.710	1.878	2.100	2.392
	MRF	1.546	1.617	1.693	1.753	1.799	1.555	1.648	1.702	1.788	1.842
	MS	1.405	1.374	1.367	1.378	1.400	1.416	1.389	1.385	1.393	1.412
Z ₂	MCR	2.344	2.428	2.500	2.584	2.651	2.355	2.440	2.506	2.596	2.694
	MI	2.427	2.616	2.893	3.201	3.668	2.431	2.649	2.900	3.246	3.703
	MRF	2.389	2.506	2.633	2.734	2.803	2.403	2.540	2.635	2.750	2.845
	MS	2.182	2.142	2.138	2.160	2.199	2.196	2.161	2.160	2.178	2.210

References

- BREIMAN, L. 2001. Random forests. *Machine learning*, 45, 5-32.
- BREIMAN L, F. J., OLSHEN RA, ET AL 1984. Classification and regression trees. *Boca Raton, FL: Chapman and Hall/CRC*.
- BURGETTE, L. F. & REITER, J. P. 2010. Multiple imputation for missing data via sequential regression trees. *American journal of epidemiology*, 172, 1070-1076.
- BUUREN, S. V. & GROOTHUIS-OUDSHOORN, K. 2010. mice: Multivariate imputation by chained equations in R. *Journal of statistical software*, 1-68.
- DOOVE, L. L., VAN BUUREN, S. & DUSSELDORP, E. 2014. Recursive partitioning for missing data imputation in the presence of interaction effects. *Computational Statistics & Data Analysis*, 72, 92-104.

- FRIEDMAN, J., HASTIE, T. & TIBSHIRANI, R. 2001. *The elements of statistical learning*, Springer series in statistics New York.
- GELFAND, A. E. & SMITH, A. F. 1990. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85, 398-409.
- HAYES, T., USAMI, S., JACOBUCCI, R. & MCARDLE, J. J. 2015. Using Classification and Regression Trees (CART) and random forests to analyze attrition: Results from two simulations. *Psychology and aging*, 30, 911.
- JAVADI, S., BAHRAMPOUR, A., SABER, M. M., GARRUSI, B. & BANESHI, M. R. 2021. 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. *Journal of Probability and Statistics*, 2021.
- LEE, K. J. & CARLIN, J. B. 2010. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *American journal of epidemiology*, 171, 624-632.
- LIAW, A. & WIENER, M. 2002. Classification and regression by randomForest. *R news*, 2, 18-22.
- MISTLER, S. A. & ENDERS, C. K. 2017. A comparison of joint model and fully conditional specification imputation for multilevel missing data. *Journal of Educational and Behavioral Statistics*, 42, 432-466.
- MORRIS, T. P., WHITE, I. R. & ROYSTON, P. 2014. Tuning multiple imputation by predictive mean matching and local residual draws. *BMC medical research methodology*, 14, 1-13.
- RAGHUNATHAN, T. E., LEPKOWSKI, J. M., VAN HOEWYK, J. & SOLENBERGER, P. 2001. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey methodology*, 27, 85-96.
- RUBIN, D. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons Inc.
- RUBIN, D. B. 2004. *Multiple imputation for nonresponse in surveys*, John Wiley & Sons.
- SEAMAN, S. R., BARTLETT, J. W. & WHITE, I. R. 2012a. Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods. *BMC medical research methodology*, 12, 46.
- SEAMAN, S. R., BARTLETT, J. W. & WHITE, I. R. 2012b. Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods. *BMC medical research methodology*, 12, 1-13.
- SHAH, A. D., BARTLETT, J. W., CARPENTER, J., NICHOLAS, O. & HEMINGWAY, H. 2014. Comparison of random forest and parametric imputation models for imputing missing data using MICE: a CALIBER study. *American journal of epidemiology*, 179, 764-774.
- SLADE, E. & NAYLOR, M. G. 2020. A fair comparison of tree-based and parametric methods in multiple imputation by chained equations. *Statistics in Medicine*, 39, 1156-1166.
- STERNE, J. A., WHITE, I. R., CARLIN, J. B., SPRATT, M., ROYSTON, P., KENWARD, M. G., WOOD, A. M. & CARPENTER, J. R. 2009. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj*, 338.
- STROBL, C., MALLEY, J. & TUTZ, G. 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological methods*, 14, 323.

- TEAM, R. C. 2013. R: A language and environment for statistical computing.
- THERNEAU, T., ATKINSON, B. & RIPLEY, B. 2018. rpart: Recursive Partitioning and Regression Trees. R package.
- VAN BUUREN, S. 2007. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical methods in medical research*, 16, 219-242.
- VAN BUUREN, S. 2018. *Flexible imputation of missing data*, CRC press.
- VON HIPPEL, P. T. 2009. 8. How to impute interactions, squares, and other transformed variables. *Sociological methodology*, 39, 265-291.
- WHITE, I. R., ROYSTON, P. & WOOD, A. M. 2011. Multiple imputation using chained equations: issues and guidance for practice. *Statistics in medicine*, 30, 377-399.