Multiply Robust Imputation for Missing Categorical Data: The Impact of Weighting Methods

Teinonen, Jonna

Multiply Robust Imputation for Missing Categorical Data: The Impact of Weighting Methods

Jonna Maria Teinonen

CBS Thesis advisors: Dr. Sander Scholtus and Prof. Dr. Ton de Waal
University Thesis advisor: Prof. Dr. Mark de Rooij

Defended on August 17, 2023

MASTER'S THESIS

STATISTICS AND DATA SCIENCE
UNIVERSITEIT LEIDEN

Acknowledgements

I would like to sincerely thank my supervisors Sander Scholtus, Ton de Waal and Mark de Rooij who made this project possible. Their invaluable guidance and continuous support carried me through all the stages of my thesis project. I would also like to express my gratitude for the staff of the CBS methodology department and for my fellow students at CBS for their support and encouragement throughout the project. Finally, I want to deeply thank my husband Pauli Teinonen. Without his unwavering love and support throughout my years of studying and this thesis project none of this would have been possible.

Contents

Abstract

Missing data is a common problem in survey research which leads to several problems, e.g., increased survey costs and biased survey estimates. Different multiple imputation (MI) methods have been developed to handle missing categorical data. One specific subset of the MI methods used for the task are the so-called robust methods, which use one or several outcome and response models to improve robustness against model misspecification. One of the robust methods, Multiply Robust Nearest Neighbour Multiple Imputation (MRNNMI), is a donor-based method that uses several outcome and response models. In MRNNMI predictive scores are obtained from all the models, weighted by using prespecified equal weights and the predictive scores are used to compute the distances between units with missing values and possible donors. In this thesis, I developed and tested the derived method Multiply Robust Imputation for Categorical Data (MRIC) which uses model quality measures to weight the predictive scores. MRIC applies the same steps as MRNNMI, but the prespecified weights are replaced by weights based on three model quality measures: four types of pseudo-R^2, the Hosmer-Lemeshow test statistic and Akaike weights. The performance of MRIC using the three different weighting approaches was compared to the existing robust MI methods that use prespecified weights, and the well-known MI approach Multivariate Imputation by Chained Equations (MICE), in a simulation study with different sample sizes and response rates. Based on the results, none of the weighting approaches influenced the imputation performance on categorical data. MRIC performed similarly to all the existing robust methods under all the conditions tested. The results indicate that for small sample sizes combined with low response rates, all the robust methods provide similar but more accurate results than MICE. However, with larger sample sizes, MICE, especially without explicit model specification, outperformed the robust methods in terms of bias and precision. Future research is needed to examine the influence of weighting based on model quality in other existing robust methods and to implement other existing model quality measures to be used in weighting the predictive scores.

Keywords: Item nonresponse, missing categorical data, multiple imputation, multiply robust imputation, model quality, goodness-of-fit

Chapter 1

Introduction

Missing data and nonresponse are common problems in survey research and several studies have demonstrated a universal trend of declining response rates in survey research (Beullens et al., 2018; de Leeuw et al., 2018). Nonresponse can be further divided into unit nonresponse, which occurs where all the data for the participant is missing, and item nonresponse where only a part of the data is missing (Graham, 2012; Schafer & Graham, 2002). Nonresponse can lead to increased survey costs, but it can also cause bias in survey estimates, variances and associations (Peytchev, 2012). These problems have increased the need to develop statistical methods to handle nonresponse. According to Little and Rubin (2020) a commonly used method for unit nonresponse is a complete-case analysis (CCA) in which the units with missing observations are completely removed from the analysis and weighting adjustments can be applied to further reduce bias. For item nonresponse, a commonly applied method is imputation which aims to fill the missing values with potential outcomes.

Imputation methods fall under two categories: single imputation (SI) and multiple imputation (MI). As the names suggest, in SI one value is imputed for each missing value, whereas in MI several values are imputed for each missing value (Little & Rubin, 2020). Although single imputation often requires less computation, the disadvantage is that it treats the imputed values as true without taking into account the uncertainty of the predictions for the missing value, unless additional steps are taken at the analysis stage. In multiple imputation, originally proposed by Rubin (1987a, 1987b), several complete data sets are created. For each data set a statistical method is applied to estimate parameters of interest and the resulting estimates are combined into pooled estimates using Rubin's rules.

Although MI is more commonly used for numerical data, several approaches for missing categorical data have been proposed. For example, MI using a latent class model (Stavseth et al., 2019; van der Palm et al., 2012; Vermunt et al., 2008) and chained equations (Akande et al., 2017; Stavseth et al., 2019; van der Palm et al., 2012) have been studied. One of the most popular ones is *Multivariate Imputation by Chained Equations (MICE)* in which each variable with missing values is regressed upon the other variables that are observed in the data (Van Buuren, 2018; Van Buuren & Groothuis-Oudshoorn, 2011). This way, it is possible to apply both linear or logistic regressions depending on the type of the missing variable. However, the disadvantage of these methods is that they only apply one working model, namely an outcome model, that predicts the missing outcome. If the working model is misspecified, e.g. a wrong link function or incorrect predictors are used, this can increase the bias of the estimates compared to more robust methods.

So called doubly robust and multiply robust imputation methods have been proposed as solutions to the problem of model misspecification. According to Chen and Haziza (2023), doubly robust methods use two working models, the outcome model and the response model, to improve the robustness of the estimates of the missing data. The former model describes the relationship between the outcome variable and the explanatory variables. The latter describes the relationship between the response indicator, a variable indicating if the data is missing, and the explanatory variables. If at least one of these working models is correctly defined, the performance of doubly robust methods is expected to remain consistent.

One example of a doubly robust method is *Doubly Robust Nearest Neighbour-based Multiple Imputation (DRNNMI)* by Zhou et al. (2017) who used the method to impute missing categorical data. The outcome model was used to predict the missing categorical outcome by using multinomial logistic regression, and the response model was used to predict the probability of the categorical outcome being missing by using logistic regression. In their study, the imputed values for each missing categorical outcome were drawn from the values of possible donors that were closest to the unit with missing values. The distance between the unit with missing values and chosen donor was calculated with the Euclidean distance by using the predicted outcome probabilities obtained from the outcome model and predictive response probabilities, also called propensity values, obtained from the response model. According to Zhou et al. (2017), if at least one of these working models is correctly specified and the response probabilities are not extreme, the estimates produced by DRNNMI should be unbiased.

Whereas doubly robust methods, like DRNNMI, use two working models to improve robustness of MI, multiply robust imputation methods use several outcome and possibly several response models as well to improve the robustness (Chen & Haziza, 2023). Chen et al. (2021) developed a *Multiply Robust Predictive Mean Matching Imputation (MRPMMI)* in which multiple outcome models are specified. In their method, the predictive values of several outcome models are used as explanatory variables in a separate regression model to predict the outcome variable. The predictive values of this regression model are used to calculate the similarity between the unit with missing values and other units with observed values. According to Chen et al. (2021) if one of the outcome models is correctly specified, the estimates should be robust against incorrect model specification. Their results showed that when incorrect outcome models were used MRPMMI outperformed estimators based on linear regression imputation and predictive mean matching with only one outcome model in terms of bias and measurement error.

Although Chen et al. (2021) applied MRPMMI to numerical data, the approach has been implemented for missing categorical data. Breemer (2022) developed a method called *Multiply Robust Nearest Neighbour Multiple Imputation (MRNNMI)* to estimate missing values for categorical data. In the method several working models are specified, as was done by Chen et al. (2021) but instead of defining only multiple outcome models, several outcome and response models are specified. The predictive values from all outcome models are computed and used as explanatory variables in a separate regression model that predicts the categorical outcome. The same is repeated for all the response models: predictive response values from all the response models are computed and used as predictors in another regression model to predict the response. The weighted sum of the predictive outcome and response values from the two separate regression models is used to calculate the distance between units with missing values and possible donors. As MRNNMI uses both working models and applies a similar Nearest Neighbour-based Multiple Imputation (NNMI) approach as was done by Zhou et al. (2017), it is considered an extension of DRNNMI. Breemer (2022) compared the performance of MRNNMI to DRNNMI and MICE in a simulation study and the results indicated that although MRNNMI was relatively robust, as it performed equally well as DRNNMI, MICE slightly outperformed the two methods in terms of bias.

Although the results by Breemer (2022) do not indicate that the use of MRNNMI would bring additional benefits over existing MI methods, it is possible that the performance of MRNNMI could be improved by weighting the predictive values based on the quality of the working models. Zhou et al. (2017) examined the optimal weighting scheme for the predictive outcome and response values for missing categorical data. Although they were unable to specify one optimal way to to define the weights, they came to the following conclusion: DRNMMI performed the best when the weights for predictive values were positive, especially if the response model was misspecified, and as long as the weights for the response values were nonnegative. The finding of nonnegative weights being optimal has been supported by other studies that used two working models in NNMI for numerical data (Hsu et al., 2014; Long et al., 2012). Based on these findings, Breemer (2022) used prespecified nonnegative weights, which were fixed and equal, in MRNNMI for the predictive outcome and response values to calculate the distance between the units with missing values and donors. However, according to Zhou et al. (2017), unequal and unfixed weights could be used if one believed that one working model is more likely to be valid than another. If the weights of the final predictive values were based on the quality of the working models, giving more weight to more valid models, it is possible that MRNNMI could provide more reliable estimates.

As both DRNNMI and MRNNMI apply multinomial and logistic regression to calculate the predictive outcome and response values, methods to quantify model quality in such regression models have to be used. There are several ways to measure model quality and most of the methods fall into one of two categories: measures of predictive power and goodness-of-fit statistics. The former describe how well independent variables predict the outcome variable and commonly used measures of predictive power in logistic regression are different types of pseudo-R^2 (Menard, 2002). The latter quantify how well the model fits the observed data, and commonly used goodness-of-fit measures are the Hosmer-Lemeshow test, the Pearson chi-square test and deviance (Hosmer et al., 2013). In addition, the Akaike information criterion (AIC) can be used as a measure for relative goodness-of-fit to compare several logistic regression models (Anderson, 2008; Menard, 2002).

The aim of this thesis is to develop and test multiply robust imputation for categorical data (MRIC) combining both doubly robust and multiply robust imputation approaches as was done by Breemer (2022). However, instead of using prespecified weights to calculate the similarity between the missing unit and the possible donors, the predicted outcome and response values will be weighted based on the quality of the working models. The quality of the models will be assessed by using the Hosmer-Lemeshow test, AIC and four commonly used types of pseudo-R^2. The performance of MRIC with different weighting methods to estimate the missing categorical variable will be compared to MRNNMI, DRNNMI, MRPMMI and MICE in a simulation study.

It is expected that the proposed method, MRIC with weights based on working model quality, will provide more reliable estimates for the missing outcome and thus result in smaller bias compared to MRNNMI, DRNNMI and MRPMMI in most conditions. In the case that MRIC performs better than the aforementioned MI methods, it is expected that it performs better or equally well as MICE. In addition, the sample size is expected to influence the performance of MRIC. When the sample size is small and the weighting is based on model quality measures that are sensitive to sample size, it is expected that the method results in larger bias compared to other MI methods. However, when the sample size is large, it is expected that the method results in more accurate estimates compared to other MI methods regardless which model quality measure was used for the weighting.

In Chapter 2, the previously described MI methods (MICE, DRNNMI, MRPMMI and MRNNMI) are described in more detail and the common ways to measure logistic regression model quality are presented. In Chapter 3 the proposed method is presented and the way the logistic model quality measures are applied in the proposed method are explained in depth. In Chapter 4 the simulation study is described: the study design, the statistical analyses and the results. In the final chapter, the findings of the thesis are summarised and discussed, the limitations of the study are stated and suggestions for future studies are given.

Chapter 2

Background

2.1 Missingness Mechanisms

Graham (2012) defines missingness in a dataset at the operational level as a binary variable R. In the case of item nonresponse R takes the value of 1 when the outcome variable Y is present and 0 when Y is missing. Different missingness mechanisms are processes in which another variable influences the value of R and the models that represent these mechanisms are called response models. For methods that handle missing data, including MI, it is important to understand the underlying missingness mechanism in the data as the results of MI are influenced by these mechanisms.

There are three common categories of missingness mechanisms: Missing Completely At Random (MCAR), Missing At Random (MAR) and Not Missing At Random (NMAR). According to Graham (2012) MCAR refers to a situation in which the probability of Y being missing is not related to any of the observed or unobserved variables or to Y itself. Therefore, as the data is not systematically missing, the observed values can be considered a simple random sample from the target population. In the case of MAR, the probability of Y being missing is systematically related to the observed but not the missing data. In particular, once R is conditioned on the observed variables in the data set, the missingness is not related to the value of Y itself. Therefore, the probability of being missing can be modelled. Finally, according to Graham (2012), when a variable is NMAR the causes for missingness are unknown: the probability of the value being missing depends on unobserved data.

For MI the data is usually assumed to be MCAR or MAR. When the data is MAR, the missingness is taken into account in MI by including the variables that are expected to influence the missingness of the outcome variable in the analysis (Graham, 2012; Van Buuren, 2018). In the case of NMAR, MI will result in biased estimates and therefore more complicated methods and further analyses of the data are required (Van Buuren, 2018).

2.2 Multiple Imputation Methods

In this section first the general assumptions for MI and Rubin's rules that are used to calculate the pooled parameters from the imputed datasets are presented. Then already existing MI methods, which are later applied in this thesis, are explained in detail.

According to Rubin (1987b) it is assumed in MI that the parameters of the data model and the parameters of the response model are distinct. In addition, as previously mentioned, all the MI methods discussed here assume that the data is MAR or MCAR (Rubin, 1987b). Suppose we have a finite population of size N from which a sample of size n is drawn. Let Y represent a categorical outcome variable with C categories that has missing values: Y^{obs} denotes the observed values and Y^{mis} the unobserved values. $\mathbf{X} = (X_1, X_2 \dots X_B)$ denotes a matrix of observed auxiliary variables that can be either categorical or numerical. Recall that the response indicator is denoted by R: $R = 1$ if Y is observed and $R = 0$ if Y is missing. To meet the MAR assumption, R can depend on Y^{obs} and $\mathbf{X}$ but not Y^{mis}:

$$P(R|Y^{obs}, \mathbf{X}, Y^{mis}) = P(R|Y^{obs}, \mathbf{X}) \tag{2.1}$$

The other common feature in all the MI methods that are covered in this thesis is that after all the data sets have been imputed, Rubin's rules (Rubin, 1987b) are applied to calculate the pooled parameters. Suppose we impute the data M times and therefore, we have M imputed datasets. Let Q_c be the parameter of interest, in our case the proportion of category c. According to Rubin (1987b) the pooled parameter estimate is calculated by

$$\bar{Q}_c = \frac{\sum_{m=1}^{M} \hat{Q}_{c,m}}{M} \tag{2.2}$$

where $\hat{Q}_{c,m}$ ($m = 1 \ldots M$) denotes the estimated proportion of category c at the mth imputation.

In order to calculate the pooled standard error (SE) for $\bar{Q}_c$, we first need to compute the so-called between variance and within variance. The former represents how much the estimates vary between different imputations and the latter represents the variance of the estimate within each dataset. The between variance is calculated by

$$V_B = \frac{\sum_{m=1}^{M} (\hat{Q}_{c,m} - \bar{Q}_c)^2}{M-1} \tag{2.3}$$

and within variance is calculated by

$$V_W = \frac{\sum_{m=1}^{M} SE_m^2}{M} \tag{2.4}$$

where SE_m^2 denotes the sampling variance of $\hat{Q}_{c,m}$ at the mth imputation. When V_B and V_W have been calculated, the pooled SE can be calculated by

$$SE_{Q_c} = \sqrt{V_W + (1 + \frac{1}{M})V_B} \tag{2.5}$$

Finally, the 95% confidence interval (CI) will be calculated by

$$CI_{95\%} = \bar{Q}_c \pm 1.96 \times SE_{Q_c} \tag{2.6}$$

For more general computation of CI, please refer to Rubin (1987b).

2.2.1 Multivariate Imputation by Chained Equations (MICE)

MICE, also called *Fully Conditional Specification (FCS)*, has become one of the common methods to handle missing data. According to Van Buuren (2018) MICE applies an iterative procedure and uses a chain of regression equations to impute Y. During each iteration these regression models use information from the observed data, i.e. conditional distributions, to impute the variables with missing values one by one until the algorithm converges. As Y can be modelled based on its distribution, the method can be used to impute both categorical and numerical data.

Suppose we have J outcome variables Y_j ($j = 1 \ldots J$) which we will denote as a matrix $\boldsymbol{Y}$ with J columns. Let Y_j represent the jth column of $\boldsymbol{Y}$ and Y_{-j} all the other columns of $\boldsymbol{Y}$ except j. Y_j^{obs} denotes all the observed values in the jth column and Y_j^{mis} all the unobserved values in the jth column. Finally, T denotes the number of iterations within the MICE algorithm and M the number of times the algorithm is repeated. Van Buuren (2018) describes the steps of the algorithm by the following:

Step 1: Define the imputation model $P(Y_j^{mis}|Y_j^{obs}, Y_{-j}, \mathbf{X}, R, \phi_j)$ for the variable Y_j with $j = 1, \ldots J$

Step 2: For each value of Y_j^{mis} we randomly draw a starting value Y_j^0 from Y_j^{obs}

Step 3: We repeat the next steps for $t = 1 \ldots T$ and $j = 1 \ldots J$.

Step 4: We define Y_{-j}^t as the currently complete data which does not include Y_j

Step 5: We draw an imputation model parameter $\phi_j^t \sim P(\phi_j|Y_j^{obs}, Y_{-j}^t, \mathbf{X}, R)$.

Step 6: We draw imputations $Y_j^t \sim P(Y_j^{mis}|Y_j^{obs}, Y_{-j}^t, \mathbf{X}, R, \phi_j^t)$

According to Van Buuren (2018) the number of iterations, T, can be fairly low, for example 5 or 10. The whole process results in M imputed data sets. In R the method can be applied by using the package **mice** (Van Buuren & Groothuis-Oudshoorn, 2011). After having obtained M imputed datasets Rubin's rules are used to calculate the pooled parameters (Rubin, 1987b).

2.2.2 Doubly Robust Nearest Neighbour Imputation (DRNNMI)

DRNNMI is a donor based MI method by Zhou et al. (2017) which uses a nearest-neighbour (NN) approach to impute categorical data. In the method predictive values from two working models, an outcome and a response model, are used to calculate the distance between units with Y^{mis} and possible donors. The steps of the method that are applied in this thesis are the following:

Step 1: From the original sample, a bootstrap sample of size n is drawn with replacement. If the bootstrap sample does not include units from all the categories of Y, a new sample is drawn.

Step 2: A multinomial outcome model is specified and fitted by regressing Y on $\mathbf{X}$ and the predictive outcome values for the units from the bootstrap sample with Y^{obs} are calculated. The multinomial logit model calculates the logarithmic odds between all the categories ($Y = 2 \ldots C$) and the reference category ($Y = 1$):

$$log\frac{P(Y=c)}{P(Y=1)} = \boldsymbol{\beta_c^T X} \qquad (c = 2, \ldots, C) \tag{2.7}$$

where $\boldsymbol{\beta_c}$ denotes a vector of $B + 1$ (the number of predictors and the intercept) regression coefficients for $Y = c$ versus $Y = 1$. Then the predictive outcome values $\mathbf{p_o}$ can be calculated by the following:

$$p_{o,c} = \frac{exp(\boldsymbol{\beta_c^T X})}{1 + exp(\boldsymbol{\beta_2^T X}) + \ldots + exp(\boldsymbol{\beta_C^T X})} \tag{2.8}$$

For each unit, this results in $C - 1$ predictive values which describe the distance between category c and the reference category. Then the predictive values are standardised so that the influence of each explanatory variable can be expressed in the same scale.

Step 3: A binomial response model is specified and fitted: R is regressed on $\mathbf{X}$. The predictive response values p_r for all the units in the bootstrap sample are calculated by the following:

$$p_r = \frac{exp(\boldsymbol{\beta^{*T} X})}{1 + exp(\boldsymbol{\beta^{*T} X})} \tag{2.9}$$

where $\boldsymbol{\beta^*}$ denotes the regression coefficients for the response model. This will result in one predictive response value per unit. The predictive values are standardised and saved.

Step 4: The predictive outcome and response values for all units with Y^{mis} from the original sample are computed by using the same estimated outcome and response models that were fitted in Steps 2-3 by using the bootstrap sample. The predictive values are standardised and saved.

Step 5: The standardised predictive outcome and response values that were computed during Steps 2-4 are used to calculate the distance between unit i from the original sample with Y^{mis} and unit j with Y^{obs} from the bootstrap sample by using an Euclidean distance function. The standardised predictive values are denoted as $\mathbf{S} = (s_1, \ldots, s_C)$ in which $s_1, \ldots s_{C-1}$ represent the $C - 1$ standardised outcome values and s_C denotes the standardised predictive response value. The distance between unit i and unit j is calculated by:

$$d(i,j) = \sqrt{\omega_1[s_{1(i)} - s_{1(j)}]^2 + \ldots + \omega_C[s_{C(i)} - s_{C(j)}]^2} \tag{2.10}$$

where $\omega_1, \ldots, \omega_C$ denote nonnegative weights that add up to 1. The weights are set to $\frac{1}{C}$ as is done in MRNNMI (See Chapter 2.2.4).

Step 6: For each unit i with Y^{mis} the k nearest donor candidates are chosen based on the calculated distance in Step 5. From the k possible donors, the value for Y^{mis} is drawn randomly from one donor. This is repeated for all the values with missing Y.

The process is repeated M times and the imputed datasets are used to calculate the pooled parameters by applying Rubin's rules (Rubin, 1987b).

2.2.3 Multiply Robust Predictive Mean Matching Imputation (MRPMMI)

MRPMMI is a method proposed by Chen et al. (2021) that applies predictive mean matching (PMM) which can be considered as a special case of NNMI. However, PMM is more often used for numerical data, as was done by Chen et al. (2021). In this thesis the method was used to impute categorical data and therefore, the NNMI approach that was used in the DRNMMI by Zhou et al. (2017) and in MRNNMI by Breemer (2022) (see also Section 2.2.4) was combined in MRPMMI with the PMM approach.

According to Chen et al. (2021) multiple outcome models can be specified in MRPMMI to improve the robustness of the MI method. In their method the predictive outcome values from each outcome model are saved and they are used as explanatory variables in a separate regression model in which Y is regressed on the saved predictive values. The latter regression will result in the final predictive values which are used to calculate the distance between unit i with Y^{mis} and unit j with Y^{obs}. The steps of the MRPMMI that are applied in this thesis are the following:

Step 1: From the original sample, a bootstrap sample of size n is drawn with replacement. If the bootstrap sample does not include units from all the categories of Y, a new sample is drawn.

Step 2: The predictive outcome values for all units from the bootstrap sample with Y^{obs} are calculated. First, V multinomial outcome regression models are specified and fitted by using the bootstrap sample. For each multinomial outcome model Y is regressed on $\mathbf{X}$ using the logarithmic odds between $C-1$ categories of Y and the reference category, similar to DRNNMI. This will result in $C-1$ predictive outcome values from each model and therefore each unit will have $V \times (C-1)$ predictive outcome values in total. Next, the final multinomial outcome regression model is fitted by regressing Y on the $V \times (C-1)$ predictive values from all the models. This will result in $C-1$ new predictive values from the final regression model that are standardised and saved.

Step 3: The predictive outcome values for units with Y^{mis} from the original sample are calculated by using the same V estimated outcome models as in Step 2 that were applied to the bootstrap sample. This will again result in $V \times (C-1)$ predictive values per unit in total. These predictive values are used in a separate regression model, which applies the same estimated final outcome model as in Step 2, as explanatory variables. The predictive values from the final regression model are standardised and saved for later analysis.

Step 4: The standardised predictive outcome values that were computed in Steps 2-3 are used to calculate the distance between unit i with Y^{mis} from the original sample and unit j with Y^{obs} from the bootstrap sample by using the Euclidean distance function. The standardised predictive outcome values are denoted as $\mathbf{S_o} = (s_1, \ldots, s_{C-1})$. The distance is calculated by the following:

$$d(i,j) = \sqrt{\omega_1[s_{1(i)} - s_{1(j)}]^2 + \ldots + \omega_{C-1}[s_{C-1(i)} - s_{C-1(j)}]^2} \tag{2.11}$$

where $\omega_1, \ldots, \omega_{C-1}$ denote nonnegative weights that add up to 1. In the thesis, the weights are set to $\frac{1}{(C-1)}$.

Step 5: For each unit i with Y^{mis} the k nearest donor candidates are chosen based on the calculated distance and the value for Y^{mis} is drawn randomly from one of the donors. The process is repeated for all the values with missing Y.

The whole process is repeated M times and Rubin's rules (Rubin, 1987b) are used to calculate the pooled parameters.

2.2.4 Multiply Robust Nearest Neighbour Multiple Imputation (MRNNMI)

MRNNMI, which was developed by Breemer (2022), combines the two approaches used in DRNNMI and MRPMMI. Namely, in MRNNMI multiple outcome and response models are specified and predictive outcome and response values are calculated in a similar manner as in NNMI. The predictive outcome and response values are then used in separate regression models as predictors as was done in PMM. The predictive values from the latter regressions are then used to calculate the distances between units with Y^{mis} and possible donors. The steps described by Breemer (2022), which are applied in this thesis, are the following:

Step 1: From the original sample, a bootstrap sample of size n is drawn with replacement. If the bootstrap sample does not include units from all the categories of Y, a new sample is drawn.

Step 2: The predictive outcome values for all the units with Y^{obs} from the bootstrap sample are calculated. We first fit V multinomial outcome models, Y is regressed on $\mathbf{X}$, by using the bootstrap sample. This will result in $V \times (C-1)$ predictive outcome values in total per unit. Next, the final multinomial regression model is fitted by regressing Y on the $V \times (C-1)$ predictive values from all the models. Correspondingly we will have $C-1$ new predictive values for each unit. The final predictive outcome values are standardised and saved.

Step 3: The predictive response values for all the units from the bootstrap sample are calculated. First W binomial response models are fitted by regressing R on $\mathbf{X}$. This will result in W predictive response values per unit. Next the final binomial response model is fitted by regressing R on the W predictive response values. The final predictive response values from this regression are standardised and saved. Correspondingly we will have one predictive response value per unit.

Step 4: The predictive outcome and response values for the units with Y^{mis} from the original sample are calculated by using the same V estimated outcome models as in Step 2 and the W estimated response models as in Step 3. The predictive values from the outcome models are again used as explanatory variables in a separate regression model that applies the same estimated final outcome model as in Step 2. The final predictive values from this regression model are standardised and saved. The same is repeated with the response models: the predictive values from each model are used as predictors and the same estimated final response model that was fitted in Step 3 is applied. The resulted predictive response values are standardised and saved. This will again result in $C-1$ predictive outcome values and one response value per unit with Y^{mis}.

Step 5: The standardised predictive outcome and response values that were computed during Steps 2-4 are then used to calculate the distance between unit i with Y^{mis} from the original sample and unit j with Y^{obs} from the bootstrap sample by using the Euclidean distance function. The standardised predictive outcome and response values are denoted as $\mathbf{S} = (s_1, \ldots, s_C)$ in which $s_1, \ldots s_{C-1}$ represent the $C-1$ standardised outcome values and s_C denotes the standardised response value. The distance is calculated by

$$d(i,j) = \sqrt{\omega_1[s_{1(i)} - s_{1(j)}]^2 + \ldots + \omega_C[s_{C(i)} - s_{C(j)}]^2} \tag{2.12}$$

where $\omega_1, \ldots, \omega_C$ denote nonnegative weights that add up to 1. In this thesis, the weights are set to $\frac{1}{C}$ as was done by Breemer (2022).

Step 6: For each unit i with Y^{mis} the k nearest donor candidates are chosen based on the calculated distance and the value for Y^{mis} is drawn randomly from one of the donors. The process is repeated for all the values with missing Y.

The whole process is repeated M times and the pooled parameters are calculated from the imputed datasets by using Rubin's rules (Rubin, 1987b). For a more detailed explanation of MRNNMI, please refer to Breemer (2022).

2.3 Model Quality Measures for Logistic Regressions

In this section some of the most common model quality measures for logistic regression are introduced and their application in the proposed method is explained in further detail in Chapter 3. The aim of the thesis is to develop and test a MI method that uses the model quality measure of the working models to weight the final predictive values instead of using prespecified weights $\omega_i = \frac{1}{C}, (i = 1, \ldots, C)$. As multinomial logistic regression is used for the outcome models and binomial regression is used for the response models, only model quality measures that are suitable for at least one of these types of regressions are considered.

As previously stated, most model quality methods either measure the predictive power of the model or the goodness-of-fit. For the predictive power, as previous studies have recommended the use of nonnegative weights in NNMI (Hsu et al., 2014; Long et al., 2012; Zhou et al., 2017), the scope in this thesis is limited to four different types of pseudo-R^2 which do not produce negative values: McFadden (1974), McKelvey and Zavoina (1975), Cox and Snell (1989) and Nagelkerke (1991). For the goodness-of-fit, only the Hosmer-Lemeshow test and the Akaike Information Criterion, AIC, will be considered. Other possible goodness-of-fit tests for logistic regression, such as the Pearson chi-square test and deviance, compare the number of observed cases to the number of expected cases which requires that the data can be grouped prior to the test (Menard, 2002). This can be especially difficult if the logistic regression model includes continuous predictors. However, in the Hosmer-Lemeshow test the data is grouped based on the predictive values of the outcome variable (Menard, 2002) and therefore, the data does not have to be grouped before applying the test. In other words, the Hosmer-Lemeshow test allows more flexibility regarding the format of the data compared to the Pearson chi-square test or deviance. For AIC, several models are needed to compare the relative goodness-of-fit, but the data does not have to be grouped.

2.3.1 Pseudo-R^2

One of the most common ways to assess the predictive power of a linear regression model is to compute the ordinary least squares (OLS) R^2 and adjusted R^2 which quantify the proportion of the variance in the target variable that is explained by the explanatory variables (Anderson, 2008; Menard, 2002). Several types of pseudo-R^2 have been proposed as an alternative for evaluating logistic regression models (Cox & Snell, 1989; McFadden, 1974; McKelvey & Zavoina, 1975; Nagelkerke, 1991). According to Hemmert et al. (2018) most of the studies conducted before 2000 that reported a measure of model fit for logistic or probit regression used pseudo-R^2. However, over 80% of these studies did not report what type of pseudo-R^2 they used (Hoetker, 2007) and currently there is no consensus in literature on which of the pseudo-R^2 methods is superior (Hemmert et al., 2018). Therefore, for simplicity, the scope in this thesis is limited to four pseudo-R^2 methods which take nonnegative values and are also commonly available in most statistical software. These pseudo-R^2, their formulas and ranges are reported in Table 1.

Table 1
The Formulas and Ranges of Four Common Pseudo-R^2 Metrics

Pseudo-R^2	Formula	Range
McFadden (1974)	$R^2_{MF} = 1 - \frac{log(L_M)}{log(L_0)}$	$0 \leq R^2_{MF} < 1$
Cox and Snell (1989)	$R^2_{CS} = 1 - exp(-\frac{2 \cdot (log(L_M) - log(L_0))}{n})$	$0 \leq R^2_{CS} < 1 - exp(\frac{2 \cdot log(L_0)}{n})$
McKelvey and Zavoina (1975)	$R^2_{MZ} = \frac{Var(\hat{y^*})}{Var(\hat{y^*}) + Var(\epsilon)}$	$0 \leq R^2_{MZ} < 1$
Nagelkerke (1991)	$R^2_N = \frac{1 - exp(-\frac{2 \cdot (log(L_M) - log(L_0))}{n})}{1 - exp(\frac{2 \cdot log(L_0)}{n})}$	$0 \leq R^2_N < 1$

Note. L_M and L_0 refer to the likelihood of the models with and without predictors. n refers to the number of observations in the model. $Var(\hat{y^*})$ is the variance of the predicted response variable. $Var(\epsilon)$ is the residual variance, which in the case of the logit link function corresponds to $\pi^2/3$.

In contrast to OLS-R^2, pseudo-R^2 values do not have a meaning that is easily or intuitively interpreted and therefore the model fit is commonly assessed by comparing the reported pseudo-R^2 to an existing benchmark value (Hemmert et al., 2018). However, as the calculations of different pseudo-R^2 values vary and studies have shown that when different pseudo-R^2 values are compared they result in very different averages (Hemmert et al., 2018; Smith & McKenna, 2013), the benchmark values also vary. In addition, pseudo-R^2 values are also influenced by the sample size, the number of categories in the outcome variables, the distribution of the outcome variable and the number of predictors (Hemmert et al., 2018).

Studies have shown that generally pseudo-R^2 methods tend to give different values compared to OLS-R^2 (DeMaris, 2002; Smith & McKenna, 2013). According to McFadden (1979), values between .2 to .4 are considered an indication of good model fit and values above the range as excellent model fit for R^2_{MF}. Hemmert et al. (2018) examined the possible benchmarks for several pseudo-R^2 values when using binomial regression models. For R^2_{CS}, with a sample size of 200 and a distribution of Y that is not highly skewed, values between .32-.58 indicate a good model fit and values above the threshold indicate an excellent fit. In contrast, when the sample size is above 200, results in the range between .25-.48 indicate a good model fit. For R^2_N the corresponding range for sample sizes smaller than 200 is .24-.44 and above 200 .17-.39. However, the study did not include R^2_{MZ}, and it is unclear whether these benchmark ranges for R^2_{MZ} have been determined in literature. Notwithstanding, several studies indicate that R^2_{MZ} is the most similar to OLS-R^2 compared to the other pseudo-R^2 values assayed (DeMaris, 2002; Veall & Zimmermann, 1996; Windmeijer, 1995).

2.3.2 The Hosmer-Lemeshow Test

Absolute goodness-of-fit measures for logistic regressions, like the Pearson Chi-square test and deviance, measure the fit of a model by comparing the fitted values with the observed ones (Hosmer et al., 2013). However, as previously mentioned, these methods require the data to be aggregated into groups, preferably with 5 or more observations in each group, or combining groups, to produce reliable measures (Hosmer et al., 2013; Menard, 2002). This prior aggregation of the data can be difficult, especially with continuous predictors. The Hosmer-Lemeshow test was developed to overcome the challenges of the previously mentioned goodness-of-fit measures.

According to Hosmer et al. (2013) in the Hosmer-Lemeshow test the data is grouped based on the estimated probabilities of the binary outcome variable. For example, by dividing the data into a fixed number of groups e.g. $G = 10$, each group g will have $n/10$ units, based on a certain cut-point value created by G. The first group will include units with the smallest predictions and last group will include the ones with the largest predictions. With $G = 10$, we divide the units into deciles determined by the predictions. The first decile forms the first group, the second decile the second group and so forth. The Hosmer-Lemeshow goodness-of-fit statistic is calculated by the following:

$$HL = \sum_{g=1}^{G} \frac{(O_g - E_g)^2}{E_g(1 - E_g/n_g)} \sim \chi^2_{G-2} \tag{2.13}$$

in which O_g denotes the number of observed cases with outcome 1 in group g, E_g the number of expected cases with outcome 1 in group g and n_g the total number of units in the gth group. In addition to the Hosmer-Lemeshow statistic, a p-value is computed from a chi-square distribution. If the p-value is significant and the HL statistic is large, this indicates a poor model fit.

When interpreting the Hosmer-Lemeshow test, it is good to keep in mind that the test is sensitive to the number of groups (Hosmer et al., 2013) and also to the sample size when the model fit is not perfect (Kramer & Zimmerman, 2007). According to Hosmer et al. (2013), when the number of groups is too low, the test will most often indicate a good model fit, whereas Kramer and Zimmerman (2007) showed that with small sample sizes the test often rejected even good models. In addition, the number of groups has been shown to impact the power of the test (Paul et al., 2013). According to Hosmer et al. (2013), a common choice for the number of groups is 10. Paul et al. (2013) recommend to use a G-value that is larger than 6 and preferably there should be at least 5 units per group. In R, the Hosmer-Lemeshow test can be performed by using the *hoslem.test()* method in the **ResourceSelection** package (Lele et al., 2019).

2.3.3 AIC and Akaike Weights

Whereas the previously mentioned goodness-of-fit measures (The Hosmer-Lemeshow test, Pearson chi-square test and deviance) are examples of an absolute comparison of fit, the Akaike Information Criterion, AIC, is a common measure for relative goodness-of-fit that can be used to compare regression models, including logistic regression models, regardless of whether the models are nested or not (Anderson, 2008; Menard, 2002). The method compares several models at once and assesses which model best fits the data: the model with the smallest AIC score is preferred as this results in the least loss of information while keeping the number of predictors as small as possible (Anderson, 2008). The AIC score is calculated as:

$$AIC = -2log(L) + 2K \tag{2.14}$$

where L denotes the likelihood that the model could have generated the observed outcome values and K denotes the number of parameters. It is good to keep in mind that the AIC values between models are only comparable if the models have been estimated by using the same data.

Anderson (2008) describes Akaike weights, also called model probabilities, which represent the probability of the model being the best (in the sense of AIC score) from the set of models given the data. The larger the Akaike weight is, the better the model fits the data from the given set of models. Akaike weights can be calculated by using the AIC scores (formula 2.14) of all models. The calculation will be demonstrated by using the outcome models as an example. Suppose we have V outcome models. For each model we need to calculate the difference between the AIC score of the current model and the AIC score of the model with the smallest value. Thus, the differences are calculated by:

$$\Delta_i = AIC_i - AIC_{min,V} \tag{2.15}$$

where AIC_i denotes the AIC score of outcome model i and $AIC_{min,V}$ denotes the AIC score of the model(s) with the smallest AIC score from the V models. After we have calculated the differences, we are able to calculate the Akaike weights, denoted as $A_{\omega(v)}$, for each model by the following:

$$A_{\omega(i)} = \frac{exp(-\frac{1}{2}\Delta_i)}{\sum_{v=1}^{V} exp(-\frac{1}{2}\Delta_v)} \tag{2.16}$$

As Akaike weights are nonnegative values, they can be used as weights to indicate the quality of each outcome model. However, according to Akande et al. (2017), with AIC scores and Akaike weights we need to keep in mind that, as they represent the relative goodness-of-fit, the values alone are not an indication of good or bad models. AIC scores indicate the best model given a set of models and if all the models fit the data poorly, the method will not be able to quantify this. Therefore, before using AIC scores and Akaike weights, one should use additional model quality measures (e.g. pseudo-R^2) to gain information about the fit of the models.

Chapter 3

Proposed Method

3.1 Proposed Method: Multiply Robust Imputation for Categorical Data (MRIC)

This chapter introduces multiply robust imputation for categorical data (MRIC), the proposed approach that applies the logistic regression model quality measures (see Chapter 2) to weight the predictive values from the working models. The first steps in MRIC are similar to MRNNMI. However, as different model quality measures require slightly different approaches for how they can be used in MRIC, the application of each model quality measure in MRIC is explained separately.

3.1.1 Weighting with Pseudo-R^2

The first version of MRIC applies the pseudo-R^2 to weight the predictive values and is called MRIC$_{R^2}$. Four different types of pseudo-R^2 measures that were introduced in Chapter 2.3.1, i.e. McFadden, Cox and Snell, McKelvey and Zavoina, and Nagelkerke, are used in MRIC$_{R^2}$. In corresponding order, these are denoted in this thesis by MRIC$_{R^2_{MF}}$, MRIC$_{R^2_{CS}}$, MRIC$_{R^2_{MZ}}$ and MRIC$_{R^2_{N}}$. The steps of the method are the following:

Step 1: As in other methods, a bootstrap sample of size n is drawn with replacement from the original sample. If the bootstrap sample does not include units from all the categories of Y, a new sample is drawn.

Step 2: The predictive outcome values for all the units with Y^{obs} from the bootstrap sample are calculated in a similar way as in Step 2 in MRNNMI. This results in $C-1$ predictive outcome values for each unit.

Step 3: The predictive response values for all the units from the bootstrap sample are calculated similar to Step 3 in MRNNMI. This will again result in one response value per unit.

Step 4: The predictive outcome and response values for the units with Y^{mis} from the original sample are calculated similarly to Step 4 in MRNNMI. Again, this will result in $C-1$ predictive outcome values and one response value per unit.

Step 5: The weights for the C predictive values are calculated by computing the pseudo-R^2 values separately for the outcome and response values.

(a) First, to calculate the pseudo-R^2 for the predictive outcome values, we use the same data that was used for the final outcome model in Step 2. Only the values that belong to the reference category or to the category c ($c = 2, \dots, C$) are chosen resulting in a sample size n^{c^*}. This data is changed to a binary format so that if $Y = c$ then $Y = 1$ and else $Y = 0$. Then, depending on which of the four types of pseudo-R^2 is used, the pseudo-R^2 for category c is calculated by using the corresponding formula in Table 1 in Chapter 2.3.1. The log-likelihood of the binary model is calculated by

$$Loglikelihood = \sum_{i=1}^{n^{c^*}} (y_i log(\hat{y}_i) + (1 - y_i) log(1 - \hat{y}_i)) \tag{3.1}$$

where y_i denotes the value of the ith unit (either 1 or 0) and $\hat{y}_i$ denotes the predicted probability of the ith unit belonging to category c. This results in $C - 1$ pseudo-R^2 values denoted as $\mathbf{R^2_{outcome}}$. Step (a) is repeated for each category $c = 2, \ldots, C$.

(b) Next, to calculate the pseudo-R^2 for the predictive response value, we use the same data that was used for the final response model in Step 3. Again, the log-likelihood of the model is calculated by using a formula similar to 3.1 and the corresponding pseudo-R^2 formula from Table 1 in Chapter 2.3.1. This will result in one pseudo-R^2 for the predictive response value denoted as $R^2_{response}$.

(c) Finally, the C pseudo-R^2 values need to be rescaled. First $\mathbf{R^2_{outcome}}$ values are rescaled so that they add up to one. To give more weight to the predictive response value, we calculate the following difference:

$$W_{outcome} = 1 - R^2_{response} \tag{3.2}$$

Each pseudo-R^2 in $\mathbf{R^2_{outcome}}$ is multiplied by $W_{outcome}$ so that the sum of $\mathbf{R^2_{outcome}}$ adds up to $W_{outcome}$. This results in C weights for the predictive outcome and response value, denoted as $\Omega = \omega_1, \ldots, \omega_C$, that add up to one. $\omega_1, \ldots, \omega_{C-1}$ represent the final weights for the outcome predictive values (rescaled $\mathbf{R^2_{outcome}}$ that was multiplied by $W_{outcome}$) and ω_C represents the final weight for the predictive response value ($R^2_{response}$).

Step 6: Similar to Step 5 in MRNNMI, the standardised predictive outcome and response values that were computed in Steps 2-4 are used to calculate the distance between unit i with Y^{mis} from the original sample and unit j with Y^{obs} from the bootstrap sample by using the Euclidean distance function. The function is the following:

$$d(i, j) = \sqrt{\omega_1 [s_{1(i)} - s_{1(j)}]^2 + \ldots + \omega_C [s_{C(i)} - s_{C(j)}]^2} \tag{3.3}$$

where $s_1, \ldots, s_{C-1}$ denote the standardised predictive outcome values, s_C the standardised predictive response value and $\omega_1, \ldots, \omega_C$ the weights for the outcome and response values calculated in Step 5.

Step 7: For each unit i with Y^{mis} the k nearest donor candidates are chosen based on the calculated distance and the donor value for Y^{mis} is randomly drawn from the candidates. The process is repeated for all the values with missing Y.

We repeat the Steps 1-7 M times which results in M imputed datasets. The pooled parameters are calculated by using Rubin's rules (Rubin, 1987b).

3.1.2 Weighting with the Hosmer-Lemeshow Chi-Square Statistic

In the second version of MRIC the weights for the predictive values are computed by using the Hosmer-Lemeshow (HL) test statistic and the method is denoted as MRIC_{HL}. Most of the steps are similar to MRIC_{R^2} and therefore these steps will not be reiterated in this section. The steps of the method are the following:

Step 1: The Steps 1-4 that were performed in MRIC_{R^2} are repeated.

Step 5: The weights for the predictive outcome and response values are calculated by computing the HL statistics.

(a) First the HL statistics are calculated for the $C-1$ predictive outcome values. As in MRIC$_{R^2}$, we use the same data that was used in Step 2 for the final outcome model. Only the values that belong to the reference category or to the category c ($c = 2, \ldots, C$) are chosen and this data is changed to a binary format so that if $Y = c$ then $Y = 1$ and else $Y = 0$. Then a separate binomial regression is fitted for every category of c by regressing Y on $\mathbf{X}$ and the HL test statistic is calculated by using formula 2.13 in Chapter 2.3.2 ($G = 10$ by default). This results in $C - 1$ HL chi-square test statistics. Step (a) is repeated for each category $c = 2, \ldots, C$.

(b) Next the HL test statistic for the predictive response value is calculated by using the data that was used in Step 3 for the final response model. Formula 2.13 is applied again with $G = 10$. This results in one HL chi-square statistic.

(c) Finally, the C HL test statistics are rescaled so that they add up to one. First, as a larger HL statistic indicates a worse model fit, the inverse of each HL statistic is calculated. Then each inverse HL statistic is divided by the total sum of the C inverse HL statistics. This results in C weights for the predictive outcome and response value, denoted as $\Omega = \omega_1, \ldots, \omega_C$, that add up to one. $\omega_1, \ldots, \omega_{C-1}$ represent the rescaled inverse HL statistics for the predictive outcome values and ω_C represents the rescaled inverse HL statistic for the predictive response value.

Step 6: As in MRIC$_{R^2}$, the standardised predictive values (computed in Steps 2-4) are used to calculate the distance between unit i with Y^{mis} from the original sample and unit j with Y^{obs} from the bootstrap sample by using the Euclidean distance function. The function is the following:

$$d(i, j) = \sqrt{\omega_1[s_{1(i)} - s_{1(j)}]^2 + \ldots + \omega_C[s_{C(i)} - s_{C(j)}]^2} \tag{3.4}$$

where $s_1, \ldots, s_{C-1}$ denote the standardised predictive outcome values, s_C the standardised predictive response value and $\omega_1, \ldots, \omega_C$ the weights for the predictive values calculated in Step 5.

Step 7: For each unit i with Y^{mis} the k nearest donor candidates are chosen based on the calculated distance and the donor value for Y^{mis} is randomly drawn from the candidates. The process is repeated for all the values with missing Y .

The steps are repeated M times and Rubin's rules (Rubin, 1987b) are used to calculate the pooled parameters from the imputed datasets.

3.1.3 Weighting with Akaike Weights

The third version of MRIC uses Akaike weights to weight the predictive outcome and response values. These weights are calculated from the AIC scores of the outcome and response models. The method is denoted as MRIC$_{\text{AIC}}$. The main difference between MRIC$_{\text{AIC}}$, MRNNMI and other MRIC methods is that in MRIC$_{\text{AIC}}$, when the predictive outcome and response values are calculated, the predictive values from the working models are not used as predictors in a final regression to compute the final predictive values. Instead, all the individual predictive values from the working models are saved, weighted based on the Akaike weight of the working models and used in the distance functions. The detailed steps of the method are the following:

Step 1: Repeat Step 1 of the previous MRIC methods.

Step 2: The predictive outcome values for all the units with Y^{obs} from the bootstrap sample are calculated. As in Step 2 in MRNNMI, V multinomial outcome models are fitted by regressing Y on $\mathbf{X}$ which results in $V \times (C - 1)$ predictive outcome values per unit. However, contrary to the MRNNMI, these $V \times (C-1)$ predictive values are standardised and saved instead of using them in a separate regression model.

Step 3: The predictive response values for all the units from the bootstrap sample are calculated. Similar to Step 3 in MRNNMI, W binomial response models are fitted by regressing R on $\mathbf{X}$ which results in W predictive response values per unit. In contrast to MRNNMI, these W predictive values are standardised and saved.

Step 4: The predictive outcome and response values for the units with Y^{mis} from the original sample are calculated by using the same V outcome models as in Step 2 and the W response models as in Step 3. This results again in $V \times (C-1)$ predictive outcome values and W response values per unit which are standardised and saved.

Step 5: The weights for all the predictive values are computed by calculating the Akaike weights for the working models.

(a) First, we calculate the Akaike weights for the V outcome models by using formulas 2.14-2.16 in Chapter 2.3.3. This will result in V Akaike weights. However, as we need weights for all the predictive outcome values, the V Akaike weights are copied $C-1$ times so that each predictive outcome value from the same outcome model gets the same weight. Note that the V outcome Akaike weights add up to 1 and as we copied the weights $C-1$ times, all the outcome Akaike weights add up to $C-1$.

(b) Next, we calculate the Akaike weights for the W response models by using the same formulas. This results in W Akaike weights for the predictive response values that add up to 1.

Step 6: The distance between unit i with Y^{mis} from the original sample and unit j with Y^{obs} from the bootstrap sample is computed by using the predictive values calculated in Steps 2-4 and the weights calculated in Step 5 in the Euclidean distance function. The standardised predictive outcome and response values are denoted as $\mathbf{S} = (s_1, \ldots, s_T)$, $(T = V \times (C-1) + W)$ in which $s_1, \ldots, s_{T-W}$ represent the standardised outcome values and $s_{(T-W)+1}, \ldots, s_T$ denote the standardised response values. The Akaike weights are denoted as $\Omega = \omega_1, \ldots, \omega_T$ where $\omega_1, \ldots, \omega_{T-W}$ represent the Akaike weights for the predictive outcome values and $\omega_{(T-W)+1}, \ldots, \omega_T$ the Akaike weights for the predictive response values. The distance is calculated as follows:

$$d(i,j) = \sqrt{\omega_1[s_{1(i)} - s_{1(j)}]^2 + \ldots + \omega_T[s_{T(i)} - s_{T(j)}]^2} \tag{3.5}$$

Note that the weights add up to C. Although this increases the absolute distances between all units the relative distances stay the same.

Step 7: For each unit i with Y^{mis} the k nearest donor candidates are chosen based on the calculated distance and the donor value for Y^{mis} is randomly drawn from the candidates. This is repeated for all the values with missing Y.

The process is repeated M times and Rubin's rules (Rubin, 1987b) are used to calculate the pooled parameters.

Chapter 4

Simulation Study

This chapter presents a simulation study conducted to compare the performance of the proposed method and existing MI methods to impute missing categorical data. First the variations of each MI method used in the simulation study are described. After this, the study design is explained and finally the results are described in detail. The R scripts of the simulation study are available in GitHub and the link to the repository can be found in Appendix A.

4.1 Methods

If all the variations of the MI methods are considered, in total 13 MI methods were used in the simulation study, where 6 of the methods were different variations of MRIC described in Chapter 3. Three of the methods were other robust methods described in Chapter 2. Finally, four different variations of MICE with different outcome models were used. Table 2 summarises each MI variation with a short description of its weighting approach and references to the relevant sections.

Table 2.
Description of All the MI Variations Used in the Simulation Study

Method	Description	Sections
$\text{MRIC}_{R^2_{MF}}$	The method used McFadden pseudo-R^2 to weight the predictive values.	2.3.1 3.1.1
$\text{MRIC}_{R^2_{CS}}$	Cox and Snell pseudo-R^2 was used to weight the predictive values.	2.3.1 3.1.1
$\text{MRIC}_{R^2_{MZ}}$	The method applied McKelvey and Zavoina pseudo-R^2 to weight the predictive values	2.3.1 3.1.1
$\text{MRIC}_{R^2_{N}}$	The Nagelkerke pseudo-R^2 was used to weight the predictive values.	2.3.1 3.1.1
MRIC_{HL}	The inverse of the Hosmer-Lemeshow chi-square statistic was used to compute the weights for the predictive values.	2.3.2 3.1.2
MRIC_{AIC}	Akaike weights calculated from AIC scores were used to weight the predictive values.	2.3.3 3.1.3
MRNNMI	The weights for all the predictive values were $\frac{1}{C}$ where C refers to the number of categories of Y.	2.2.4
DRNNMI	The method used the same weighting approach as MRNNMI.	2.2.2
MRPMMI	The weights of the predictive outcome values were set to $\frac{1}{(C-1)}$ where C refers to the number of categories of Y.	2.2.3
$\text{MICE}_{\text{default}}$	MICE with default settings was used: the ith sample was given to the method without defining the outcome model. The method did not use any predefined weighting method.	2.2.1
MICE_{CM}	The method was given both the ith sample and the correctly specified outcome model (CM; see Table B6 in Appendix B). The method did not use any predefined weighting method.	2.2.1
MICE_{INM}	The ith sample and an incorrect nested outcome model (INM; see Table B6 in Appendix B) was given for the method. No predefined weighting method was applied.	2.2.1
$\text{MICE}_{\text{INNM}}$	The ith sample and an incorrect nonnested outcome model (INNM; see Table B6 in Appendix B) was given for the method. The method did not apply any predefined weighting method.	2.2.1

Note. Sections refers to the relevant chapter sections regarding the method. *CM* refers to the correct model, *INM* to an incorrect nested model and *INNM* to an incorrect nonnested model.

In this thesis, when referring to the robust methods, this includes the different variations of the proposed method (MRIC$_{R^2_{MF}}$, MRIC$_{R^2_{CS}}$, MRIC$_{R^2_{MZ}}$, MRIC$_{R^2_{N}}$, MRIC$_{\text{HL}}$ and MRIC$_{\text{AIC}}$) and the previously existing robust methods (MRNNMI, DRNNMI, MRPMMI) unless stated otherwise. For all the MI methods the number of imputations (M) was set to five as based on the results by Breemer (2022), increasing the number of imputations did not significantly improve MRNNMI or DRNNMI estimates. In addition, as all the robust methods apply the k-Nearest Neighbour approach, k was set to 5 in the current study as this has been the common choice in other similar studies (Breemer, 2022; Hsu et al., 2014; Long et al., 2012; Zhou et al., 2017).

4.2 Study Design

The study design partly follows the procedure applied by Breemer (2022) and Chen et al. (2021). As previously mentioned, most model quality measures for logistic regressions are influenced by the different factors in the study design, for example the sample size. The scope in this thesis was limited to study the influence of sample size, response rate and impact of different working models on the performance of MI methods. The current study has in total $5 \times 5 \times 3$ conditions: five different working model scenarios, five different sample sizes and three different response rates. In this subsection, first the data generation is explained and after this the previously mentioned conditions are described in detail.

First, four auxiliary variables were generated for the main population of size $N = 100000$: $X_1 \sim Binom(1, 0.5)$, $X_2 \sim N(4, 1)$, $X_3 \sim N(8, 2)$ and $X_4 \sim N(3, 1)$. An outcome variable Y with three categories ($c = 1, \dots, C$) was generated from a multinomial distribution by the following model using the first three generated auxiliary variables:

$$Y \sim X_1 + X_2 + X_3 \tag{4.1}$$

For the multinomial distribution model only logit link functions were used. The equations used to compute the average probabilities for the three categories of Y are given by

$$P(Y = 1) = 1 - P(Y = 2) - P(Y = 3) \tag{4.2}$$

$$P(Y = 2) = \frac{exp(-1.6 \cdot X_1 + 0.3 \cdot X_2 + 0.1 \cdot X_3)}{(1 + exp(-1.6 \cdot X_1 + 0.3 \cdot X_2 + 0.1 \cdot X_3) + exp(X_1 - 1.3 \cdot X_2 + 0.8 \cdot X_3))} \tag{4.3}$$

$$P(Y = 3) = \frac{exp(X_1 - 1.3 \cdot X_2 + 0.8 \cdot X_3)}{(1 + exp(-1.6 \cdot X_1 + 0.3 \cdot X_2 + 0.1 \cdot X_3) + exp(X_1 - 1.3 \cdot X_2 + 0.8 \cdot X_3))} \tag{4.4}$$

which resulted in average proportions of approximately 10.0% for category 1, 37.7% for category 2 and 52.3% for category 3.

After generating the population, six working models were defined: three outcome models and three response models. For the robust MI methods five different scenarios regarding the correctness of the working models were examined:

1. One outcome and one response model were correctly defined. For MRIC and MRNNMI the second outcome and response model were incorrect nested models and the third outcome and response model were incorrect nonnested models. For DRNNMI one correct outcome and response model were defined and for MRPMMI the same three outcome models as for MRIC and MRNNMI were used.

2. Only one outcome model was correctly defined. For MRIC and MRNNMI the second and third outcome models were the same incorrect outcome models as in scenario 1. All the response models were incorrect: two of the models were nested and the third model was a nonnested model. For DRNNMI one correct outcome model and one incorrect nested response model were defined. For MRPMMI the same outcome models were used as for MRIC and MRNNMI.

3. Only one response model was correctly defined. The second response model was an incorrect nested model and the third one an incorrect nonnested model. All the outcome models were incorrect: two of the models were nested and one was nonnested. For DRNNMI the correct response model and incorrect nested outcome models were used. Again for MRPMMI the same incorrect outcome models were used as for MRIC and MRNNMI.

4. All the working models were incorrect nested models. For DRNNMI one incorrect nested outcome and response model were defined. For MRPMMI again the same incorrect nested outcome models were used as for MRIC and MRNNMI.

5. All the working models were incorrect nonnested models. For DRNNMI one incorrect nonnested outcome and response model was used. MRPMMI used the same incorrect nonnested outcome models as for MRIC and MRNNMI.

The exact models used for all the robust methods in every scenario can be seen in Tables B1-B5 in Appendix B. As mentioned earlier, only one outcome model can be defined for MICE and therefore three different versions of MICE were defined: one using the correct outcome model, one with an incorrect nested and one with an incorrect nonnested model. The models used for the different MICE versions can be seen in Table B6 in Appendix B.

For each model scenario, five simple random samples of different sizes were drawn with replacement: 50, 100, 200, 500 and 1000. For each sample size, three different response rates were used: 60%, 70% and 80%. The response indicator was generated from a Bernoulli distribution by using different formulas to calculate the average response rates. To obtain an approximately 60% average response rate the following formula was used:

$$P(R=1) = \frac{exp(-19.3 + 1.5 \cdot X_1 + 2.5 \cdot X_2 + 1.2 \cdot X_3)}{1 + exp(-19.3 + 1.5 \cdot X_1 + 2.5 \cdot X_2 + 1.2 \cdot X_3)} \tag{4.5}$$

An average of 70% response rate was obtained by

$$P(R=1) = \frac{exp(-18.2 + 1.5 \cdot X_1 + 2.5 \cdot X_2 + 1.2 \cdot X_3)}{1 + exp(-18.2 + 1.5 \cdot X_1 + 2.5 \cdot X_2 + 1.2 \cdot X_3)} \tag{4.6}$$

and finally an average of 80% by

$$P(R=1) = \frac{exp(-17 + 1.5 \cdot X_1 + 2.5 \cdot X_2 + 1.2 \cdot X_3)}{1 + exp(-17 + 1.5 \cdot X_1 + 2.5 \cdot X_2 + 1.2 \cdot X_3)} \tag{4.7}$$

The influence of the explanatory variables on the outcome variable and the response indicator was studied for all the sample sizes by picking a simple random sample 300 times using the type II and type III Likelihood Ratio Tests (LRT). Both tests produced similar results for the outcome variable: as expected, X_1, X_2 and X_3 had a significant influence on Y, whereas X_4 did not have a significant impact. These results were the same for all the sample sizes and therefore, only the results for the smallest ($n = 50$) and largest sample size ($n = 1000$) are reported in Table C1 in Appendix C. For the response indicator R, both tests produced the same results: contrary to the expectations, only X_2 and X_3 had a significant influence on R with sample sizes 50 and 100. With sample sizes larger than 100, X_1, X_2 and X_3 had a significant influence on R as expected. For the smaller sample sizes, the LRT statistic of X_1 is larger than X_4, but X_2 and X_3 seem to contribute much more to R and therefore the influence of X_1 is not significant. The detailed results of the analysis can be found in Tables C2-C5 in Appendix C.

4.3 Statistical Analyses

For each combination of conditions, $N_{sample} = 5000$ samples were drawn and analysed using the 13 different MI methods. For all the 13 MI methods the pooled parameters were computed by using Rubin's rules (1987b; see Section 2.2) for each sample. The performance of the MI methods was compared by studying the average estimated proportion, the average standard error, the standard deviation and the coverage rate of each method.

The average estimated proportion (AEP) was calculated for each method by

$$AEP_c = \frac{\sum_{i=1}^{N_{sample}} \bar{Q}_{c(i)}}{N_{sample}} \tag{4.8}$$

where $\bar{Q}_c$ denotes the average estimated proportion of category c over M imputations. The average standard error (ASE) across simulations was given by

$$ASE_c = \frac{\sum_{i=1}^{N_{sample}} SE_{Q_{c(i)}}}{N_{sample}} \tag{4.9}$$

where $SE_{Q_{c(i)}}$ denotes the pooled standard error of $\bar{Q}_c$ over M imputations. The standard deviation (SD), which describes how much on average the pooled category estimates vary over the simulations, was computed by

$$SD_c = \sqrt{\frac{\sum_{i=1}^{N_{sample}} (\bar{Q}_{c(i)} - AEP_c)^2}{N_{sample} - 1}} \tag{4.10}$$

where $\bar{Q}_c$ represents the average estimated proportion of category c of the ith sample.

Finally, for each sample $CI_{95\%}$ was computed to check if the true population proportion for category c was within the interval. The coverage rate (CR), which is the percentage of times the 95% confidence interval covers the true population proportion, was calculated by taking the average of how frequently the true population proportion was included within the $CI_{95\%}$ over N_{sample}.

4.4 Results

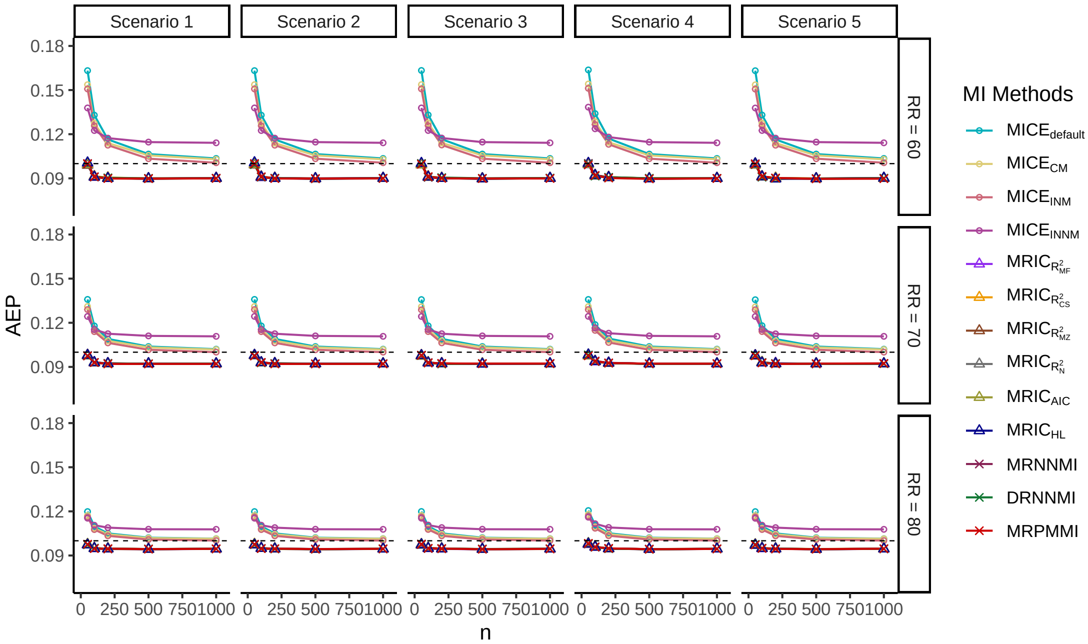
As previously mentioned, 5000 simple random samples were drawn to estimate the performance of each MI method. If the sample did not include at least one unit from all the categories of Y, a new sample was drawn. In addition, if all the observations of Y in the sample had a same predicted probability (e.g. 1), a new sample was drawn as otherwise the Hosmer-Lemeshow test failed to divide the sample into any number of groups. The latter error occurred only with sample sizes of 200 or smaller and often less than 2% of the time. Table D1 in Appendix D shows the average response rates for each sample and the percentage of how often, out of 5000, a new sample had to be drawn because the Hosmer-Lemeshow test was unable to divide the sample into at least two groups. As the average response rates are close to each other and a new sample did not have to be drawn often, this is not expected to influence the results significantly. However, for good measure, all the MICE variants were run for each scenario using the same sample as the robust methods but using the same outcome models in all the scenarios (see Section 4.2).

4.4.1 General Findings

First, the differences between the working model scenarios and the influence of the sample size and response rate on AEP are described separately for each category of Y. Figure 1 shows the AEP for category 1 (the true population proportion is 10.0%) of all the MI methods in each scenario for different sample sizes and response rates. Please note that although some of the differences may seem large in the figure, the figure's scale has been adjusted in order to highlight the relatively small differences between the MI methods.

Figure 1.

The Average Estimated Proportion for Category 1 of Y for All the MI Methods



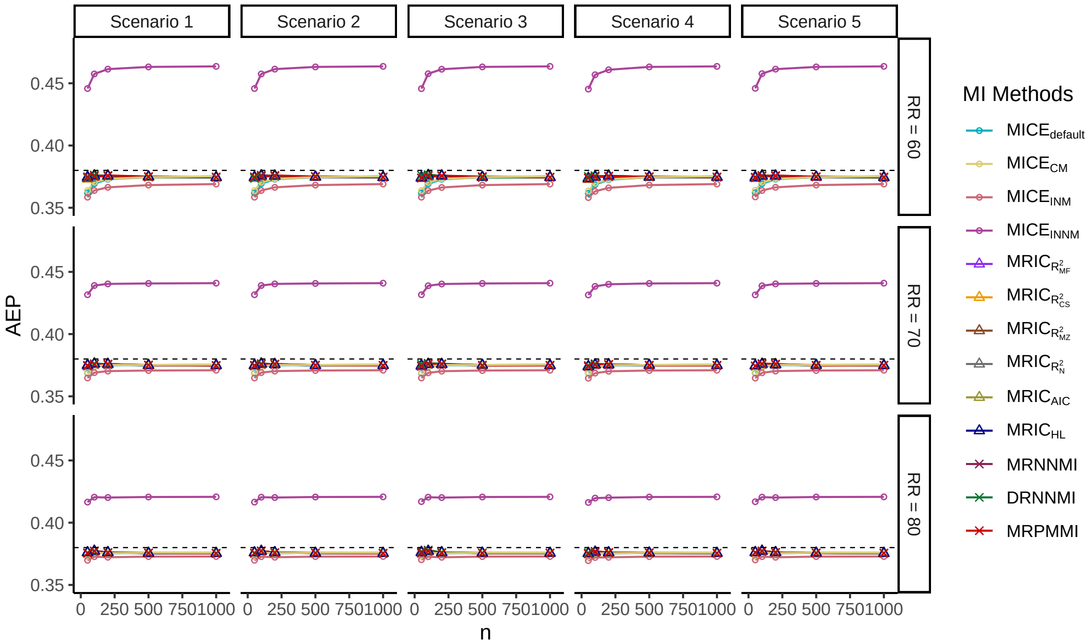
Note. Scenarios refer to the different working model scenarios that were described in Section 4.2. *AEP* refers to the average estimated proportion for category 1 ($P(Y = 1) = .100$, denoted by the dashed line), *RR* to the response rate and n to the sample size. Although the MICE variants were run for each scenario using the same sample as the corresponding robust methods in that scenario, each MICE variant used the same outcome model across all the scenarios (see Table B6 in Appendix B).

As can been seen from the figure, the different working model scenarios do not seem to have an impact on the estimates of the MI methods. The robust methods perform similarly in all the conditions and are therefore grouped together. This can be seen in Figures E1-E3 in Appendix E that illustrate the close-up results for robust methods in Scenario 2 for each category of Y. This indicates that different weighting methods do not seem to improve the results for the robust methods. For small sample sizes, robust methods seem to produce relatively unbiased, slightly underestimated, estimates whereas all the MICE variants overestimate the percentage of units belonging to category 1. As the sample size increases, the estimates of the MICE variants seem to get closer to the true population proportions, especially when the response rate is low. For the robust methods, the estimates are already close to the true values for the small sample sizes and the estimates do not seem to change or improve when the sample size is 200 or larger. The figure also demonstrates that as the response rate increases, the estimates of all the MI methods become more unbiased, even for smaller sample sizes. However, the estimates of the MICE variants seem to improve more when the response rate increases compared to the robust methods.

Figure 2 demonstrates the AEP of each MI method for category 2 (the true population proportion is 37.6%) in different model scenarios for all the sample sizes and response rates. Again, please note that the scale for the AEP has been adjusted in order to illustrate the small differences between the MI methods.

Figure 2.

The Average Estimated Proportion for Category 2 of Y for All the MI Methods



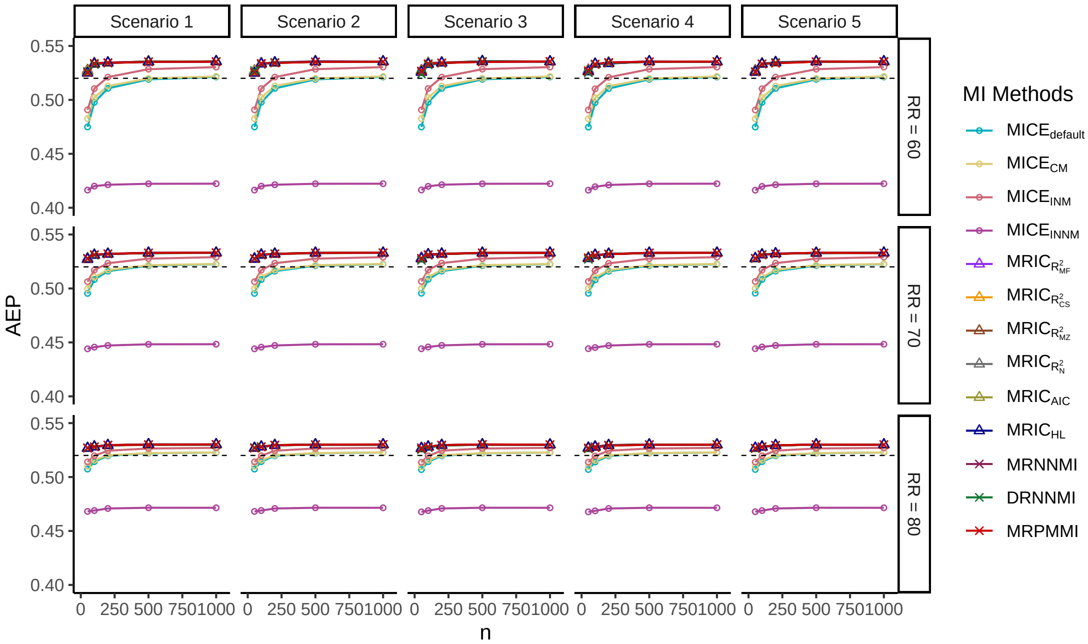
Note. Scenarios refer to the different working model scenarios that were described in section 4.2. *AEP* refers to the average estimated proportion for category 2 ($P(Y = 2) = .376$, denoted by the dashed line), *RR* to the response rate and n to the sample size. Note that although the MICE variants were run for each scenario so that they used the same sample as the robust methods in that scenario, each MICE variant used the same outcome model in all the scenarios (see Table B6 in Appendix B).

Similar to the findings of category 1, different working model scenarios do not seem to influence the performance of any of the robust methods. Again, the robust methods are all grouped together in all the conditions and the weighting does not seem to have any influence on the estimates. Overall, it can be seen that $\text{MICE}_{\text{INNM}}$ overestimates the proportion of units belonging to category 2 in all the conditions compared to other MI methods. The figure also demonstrates that as the sample size increases, the estimates of MICE variants, apart from $\text{MICE}_{\text{INNM}}$, seem to get closer to the true value, especially when the response rate is low. In contrast, increasing the sample size does not seem to have much influence on the performance of the robust methods, as the estimates are relatively unbiased even with small sample sizes and low response rates. The same pattern can be seen as the response rate increases: the estimates of all the MICE variants seem to improve as response rate increases, especially for smaller sample sizes. For robust methods, we do not see much improvement as the estimates are already quite close to the true population proportions.

Figure 3 shows the AEP of each MI method for category 3 (true population proportion is 52.3%) in each scenario for all the sample sizes and response rates. As in the previous figures, please note that the scale for AEP has been adjusted.

Figure 3.

The Average Estimated Proportion for Category 3 of Y for All the MI Methods



Note. Scenarios refer to the different working model scenarios that were described in section 4.2. *AEP* refers to the average estimated proportion for category 3 ($P(Y = 3) = .523$, denoted by the dashed line), *RR* to the response rate and n to the sample size. Note that although the MICE variants were run for each scenario so that they used the same sample as the robust methods in that scenario, each MICE variant used the same outcome model in all the scenarios (see Table B6 in Appendix B).

Similar to the previous findings, the estimates of all the robust methods are grouped and the methods perform similarly in all the conditions. The robust methods seem to slightly overestimate the number of units in the category 3 and the estimates do not improve as the sample size increases. In contrast, all the MICE variants underestimate the proportions for the small sample sizes but the estimates become less biased as the sample size increases, except for $\text{MICE}_{\text{INNM}}$. Similar to the findings in category 2, the estimates of $\text{MICE}_{\text{INNM}}$ have the largest bias of all the methods. For the response rates, it can be seen that the estimates of all the methods get closer to the true values as response rate increases but the effect seems to be larger for the MICE variants.

The results have demonstrated that the performance of the robust methods does not differ between the working model scenarios. Therefore, the rest of the analysis will be limited to Scenario 2 as it provides an example of a situation in which one of the outcome models is correctly defined but all the response models are incorrect. The comparison will be limited to the sample sizes 50 and 500 within Scenario 2 to provide an overview of how the methods perform for small and larger sample sizes. The results of all the MI methods in each scenario for all the sample sizes and response rates can be seen in Appendix F.

4.4.2 Results with Sample Size 50

Table 3 shows the AEP, ASE, SD and CR of all the MI methods for each category of Y in Scenario 2 when the sample size is 50. In addition a complete-case analysis (CCA) was performed to estimate the presence of possible bias in the response data by calculating the percentage of units in each category after removing the units with missing values from the sample.

Table 3.

The Results of the Different MI methods in Scenario 2 for $n = 50$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.096	-	.049	-	.375	-	.088	-	.529	-	.091	-
	70%	.096	-	.047	-	.375	-	.082	-	.529	-	.085	-
	80%	.096	-	.045	-	.376	-	.076	-	.528	-	.079	-
$\text{MRIC}_{R^2_{MF}}$	60%	.099	.060	.051	.923	.374	.100	.093	.941	.526	.103	.095	.941
	70%	.098	.053	.048	.908	.374	.088	.085	.940	.528	.091	.088	.941
	80%	.097	.048	.046	.912	.376	.080	.078	.941	.527	.083	.081	.942
$\text{MRIC}_{R^2_{CS}}$	60%	.100	.060	.051	.927	.374	.099	.092	.934	.527	.102	.096	.942
	70%	.098	.053	.048	.915	.374	.088	.085	.940	.528	.091	.088	.942
	80%	.098	.048	.045	.915	.376	.080	.078	.941	.527	.082	.081	.943
$\text{MRIC}_{R^2_{MZ}}$	60%	.099	.060	.050	.926	.374	.100	.092	.940	.527	.103	.095	.940
	70%	.098	.053	.048	.916	.375	.088	.085	.939	.528	.091	.088	.942
	80%	.097	.048	.046	.908	.376	.080	.078	.942	.527	.083	.081	.941
$\text{MRIC}_{R^2_{N}}$	60%	.100	.060	.051	.922	.374	.100	.092	.940	.526	.104	.096	.941
	70%	.098	.053	.048	.914	.375	.089	.085	.940	.527	.091	.089	.941
	80%	.097	.048	.046	.909	.376	.080	.078	.947	.527	.083	.081	.944
MRIC_{AIC}	60%	.100	.058	.051	.923	.374	.097	.092	.936	.527	.101	.095	.944
	70%	.098	.052	.048	.914	.375	.087	.085	.938	.527	.090	.088	.936
	80%	.098	.048	.046	.913	.376	.079	.078	.941	.527	.082	.081	.939
MRIC_{HL}	60%	.101	.068	.053	.920	.375	.112	.095	.941	.525	.116	.098	.947
	70%	.098	.058	.050	.908	.375	.097	.087	.946	.527	.100	.090	.947
	80%	.097	.050	.046	.903	.376	.084	.079	.948	.527	.087	.082	.950
MRNNMI	60%	.100	.060	.051	.930	.374	.100	.093	.937	.525	.103	.096	.943
	70%	.098	.053	.048	.916	.374	.088	.085	.938	.528	.091	.088	.940
	80%	.097	.048	.046	.911	.376	.080	.078	.944	.527	.083	.081	.942
DRNNMI	60%	.099	.058	.050	.927	.374	.096	.091	.933	.527	.099	.094	.935
	70%	.098	.051	.048	.913	.375	.086	.084	.937	.528	.089	.087	.940
	80%	.097	.047	.046	.909	.375	.079	.078	.940	.527	.081	.081	.942
MRPMMI	60%	.100	.061	.051	.928	.375	.102	.093	.938	.525	.105	.096	.939
	70%	.097	.053	.048	.914	.375	.090	.085	.943	.528	.093	.088	.943
	80%	.098	.048	.046	.912	.376	.081	.078	.942	.527	.083	.081	.943
$\text{MICE}_{\text{default}}$	60%	.163	.060	.076	.814	.362	.074	.081	.905	.475	.076	.079	.886
	70%	.136	.054	.063	.895	.369	.073	.077	.925	.495	.075	.076	.923
	80%	.120	.050	.053	.927	.373	.072	.073	.937	.507	.074	.075	.936
MICE_{CM}	60%	.154	.058	.076	.837	.364	.075	.083	.904	.483	.077	.082	.886
	70%	.131	.053	.062	.897	.369	.073	.078	.917	.500	.075	.078	.917
	80%	.117	.049	.053	.925	.373	.072	.074	.933	.510	.074	.076	.937
MICE_{INM}	60%	.151	.058	.067	.875	.358	.074	.078	.913	.491	.077	.081	.908
	70%	.129	.053	.058	.925	.365	.073	.075	.927	.506	.076	.078	.929
	80%	.116	.049	.051	.930	.370	.072	.073	.930	.514	.074	.075	.935
$\text{MICE}_{\text{INNM}}$	60%	.138	.057	.067	.888	.446	.079	.089	.818	.416	.077	.079	.686
	70%	.124	.053	.059	.915	.432	.077	.083	.873	.444	.075	.078	.791
	80%	.115	.049	.053	.920	.417	.075	.077	.909	.468	.074	.075	.872

Note. In Scenario 2 one of the outcome models is correctly defined and all the response models are incorrectly defined. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation of the estimated proportions and *CR* to the coverage rate.

The results show that the bias for each category in the response was small as the AEPs of the complete-case analysis are close to the true population proportions. As previously noted, none of the weighting methods seem to influence the bias of the estimates as the AEPs of the robust methods are close to each other. Overall, the AEPs of the robust methods for all the Y categories are less biased compared to the AEPs of the MICE variants for all the response rates. As before, the AEPs of the MICE variants get closer to the true population proportions as the response rate increases. However, increasing the response rate does not significantly improve AEPs of the robust methods as they are relatively close to the true population proportions, even when the response rate is low.

The robust methods perform similarly in terms of the ASE: the true SDs of these methods are overestimated by their ASEs, especially for lower response rates. MRIC_{HL} has the largest ASE of all the MI methods for all the Y categories. In contrast, the ASEs of the MICE variants seem to underestimate their SDs. For the larger categories (2 and 3) the MICE variants overall have lower ASEs compared to the robust methods, especially for the lower response rates. For all the MI methods, the ASEs decrease as the response rate increases, but the effect is larger for the robust methods. The SD follows a similar pattern in the robust methods: all the methods perform similarly and the SD decreases as the response rate increases. The estimates of the MICE variants seem to vary more for the smallest category as the SD of the MICE variants is larger for category 1 compared to the robust methods. However, for the larger categories, the MICE variants have smaller SDs compared to the robust methods.

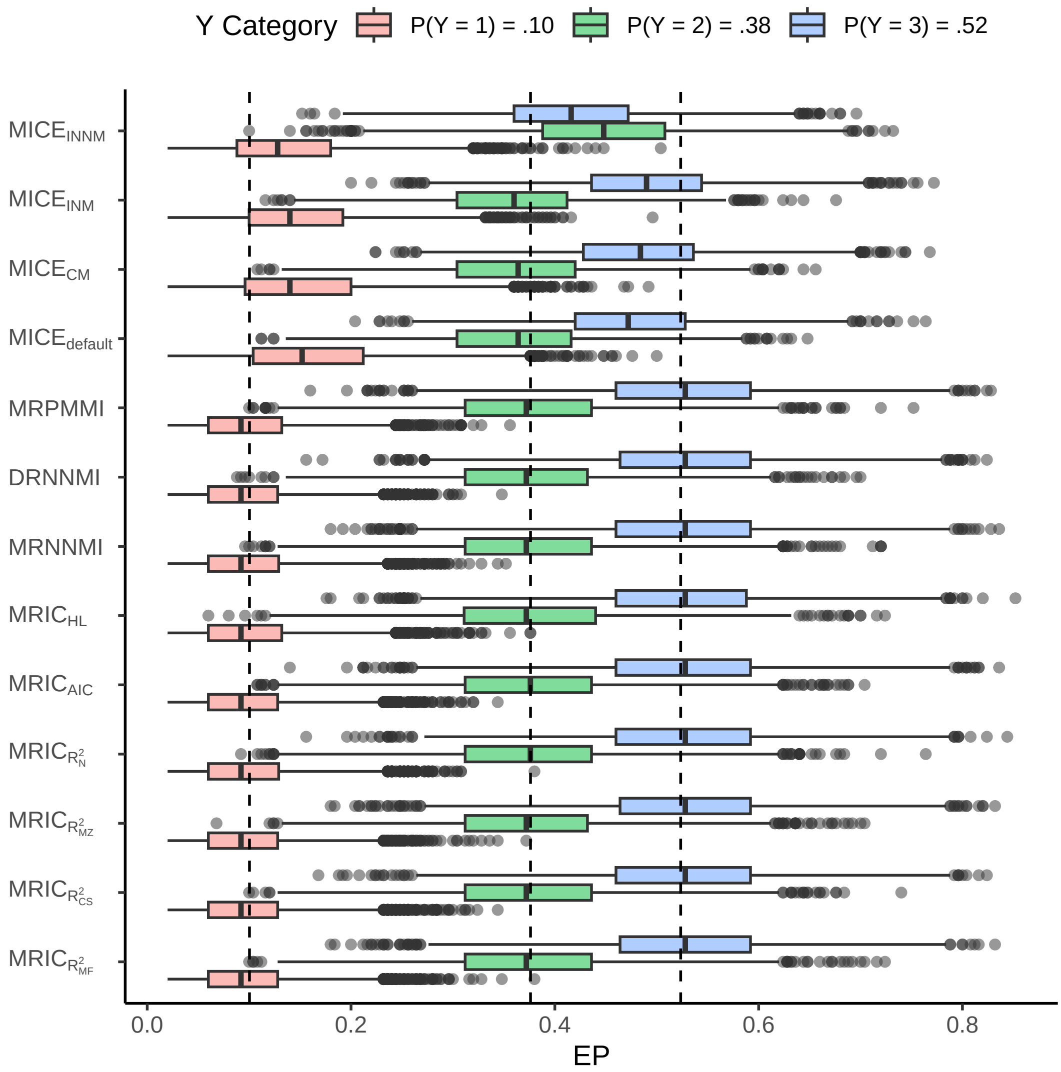
Finally, for all the MI methods, the smallest category seems to be the most challenging to estimate as the CRs are smaller for all the methods in category 1 compared to other categories. There are small but inconsistent differences in the CRs among the robust methods and the CRs for the robust methods are overall higher compared to the MICE variants. However, for the robust methods, increasing the response rate does not consistently improve the CRs, whereas for the MICE variants it does.

In summary, weighting the predictive values based on model quality did not reduce bias or improve precision in the robust methods. The robust methods all produced similar estimates that were close to the true value and less biased than the estimates of the MICE variants. Of all the methods, $\text{MICE}_{\text{INNM}}$ produces the most biased estimates, except for category 1, and has the lowest CR. However, in terms of precision, the MICE variants outperformed the robust methods for most categories as they had lower SD for categories 2 and 3. All the robust methods tend to overestimate ASEs whereas the MICE variants underestimate them.

To further illustrate the results, Figures 4-6 show the distribution of the estimated proportions of all the MI methods for the different response rates.

Figure 4.

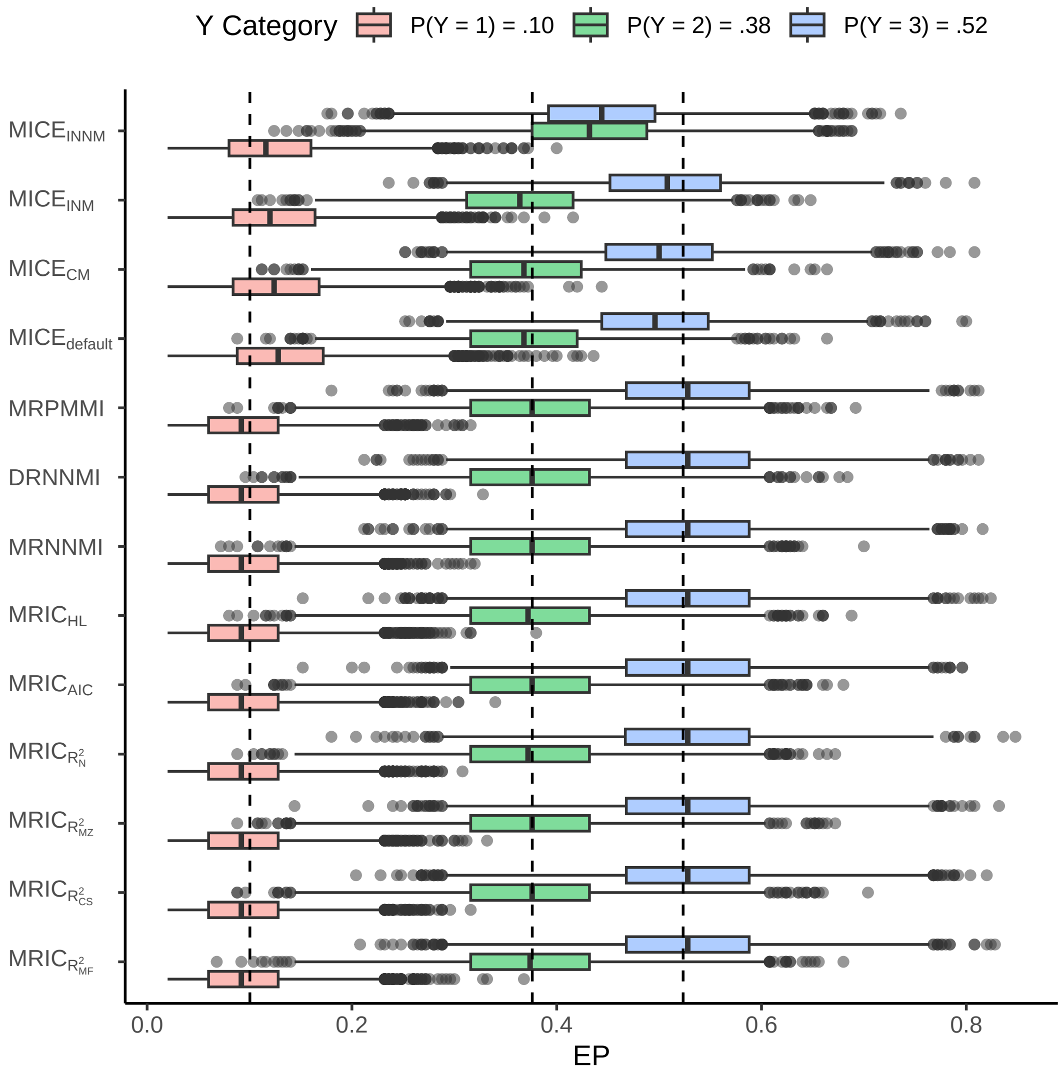
The Distributions of the Estimated Proportions for the 5000 Samples When $RR = 60\%$



Note. The figure shows the distributions of the estimated proportions in Scenario 2 for $n = 50$ when $RR = 60\%$. EP refers to the estimated proportion and RR to the response rate.

Figure 5.

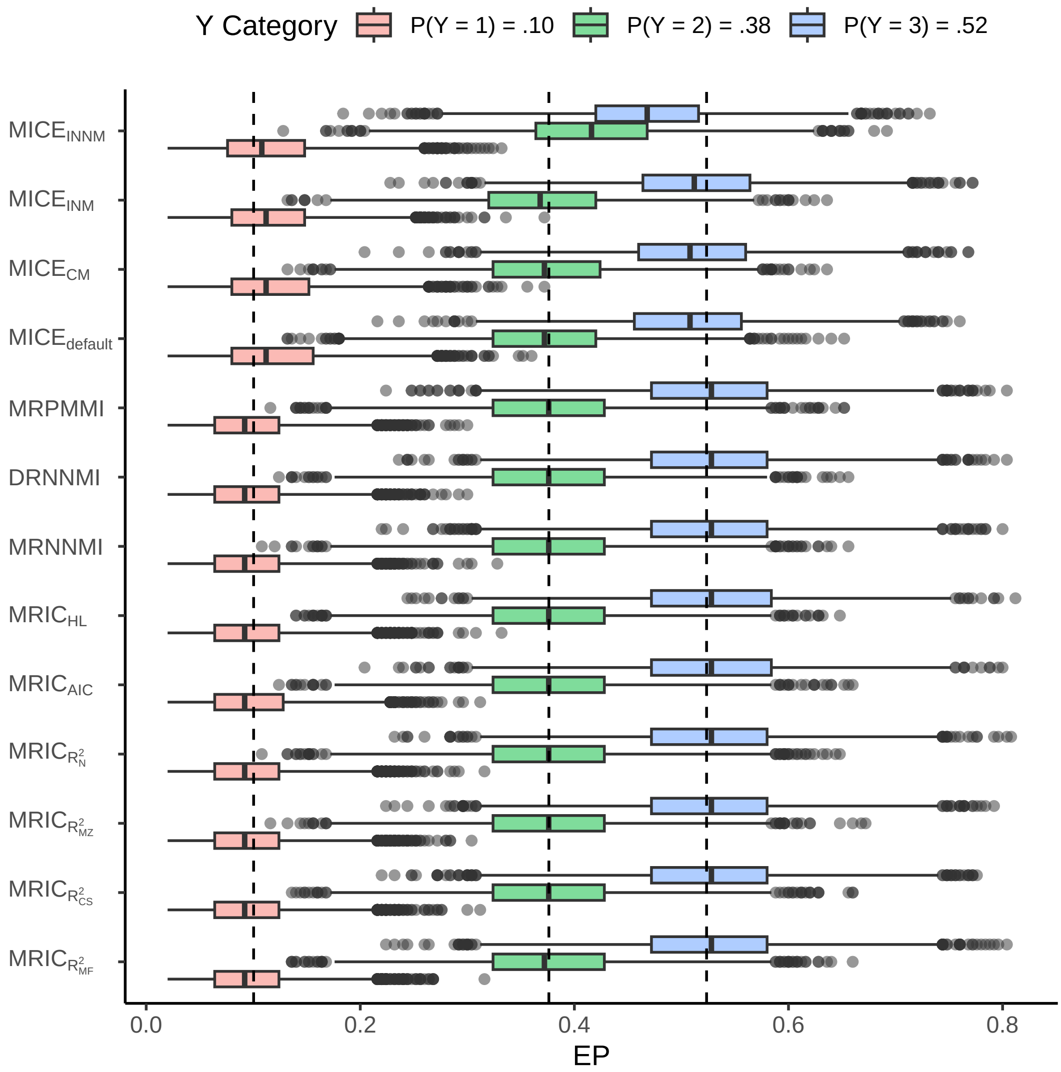
The Distributions of the Estimated Proportions for the 5000 Samples When $RR = 70\%$



Note. The figure shows the distributions of the estimated proportions in Scenario 2 for $n = 50$ when $RR = 70\%$. EP refers to the estimated proportion and RR to the response rate.

Figure 6.

The Distributions of the Estimated Proportions for the 5000 Samples When $RR = 80\%$



Note. The figure shows the distributions of the estimated proportions in Scenario 2 for $n = 50$ when $RR = 80\%$. *EP* refers to the estimated proportion and *RR* to the response rate.

The distributions of all the MI methods are relatively wide, most likely due to the small sample size which decreases the accuracy of the estimates. The figures demonstrate that the proposed methods have similar distributions as the existing robust methods for all the categories and response rates. This indicates that weighting predictive values based on model quality did not improve the performance of the robust methods. The figures show well that the estimates of all the robust methods do not change substantially as the response rate increases, but the AEPs are already close to the true population estimates for all the categories. With the exception of $\text{MICE}_{\text{INNM}}$, the improvement in the estimates for MICE variants can be seen as the response rate increases, especially for category 1. Furthermore, the distributions of the estimated proportions for category 2 and 3 of $MICE_{INNM}$ seem to overlap, indicating that the method fails to estimate the proportion of the two largest categories.

4.4.3 Results with Sample Size 500

Table 4 shows the AEP, ASE, SD and CR for all the MI methods in Scenario 2 when the sample size is 500. As with sample size 50, CCA was performed to estimate the presence of bias in the response data.

Table 4.
The Results of the Different MI methods in Scenario 2 for $n = 500$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.090	-	.016	-	.375	-	.027	-	.535	-	.029	-
	70%	.092	-	.015	-	.375	-	.026	-	.533	-	.027	-
	80%	.094	-	.014	-	.376	-	.024	-	.530	-	.025	-
$\text{MRIC}_{R^2_{MF}}$	60%	.090	.020	.017	.890	.374	.034	.029	.961	.536	.036	.030	.946
	70%	.092	.018	.016	.897	.375	.029	.027	.957	.533	.030	.028	.948
	80%	.094	.016	.015	.914	.376	.026	.025	.958	.530	.027	.026	.946
$\text{MRIC}_{R^2_{CS}}$	60%	.090	.020	.018	.890	.375	.034	.029	.959	.536	.036	.031	.947
	70%	.092	.017	.016	.905	.375	.029	.027	.951	.533	.030	.028	.942
	80%	.094	.016	.015	.915	.376	.026	.025	.954	.530	.027	.026	.946
$\text{MRIC}_{R^2_{MZ}}$	60%	.090	.020	.018	.892	.375	.035	.030	.954	.535	.036	.031	.943
	70%	.092	.017	.016	.901	.375	.029	.027	.956	.533	.030	.028	.945
	80%	.094	.016	.015	.915	.375	.026	.025	.952	.530	.027	.026	.943
$\text{MRIC}_{R^2_{N}}$	60%	.090	.020	.018	.883	.375	.034	.030	.957	.536	.036	.031	.944
	70%	.092	.018	.016	.900	.375	.029	.027	.957	.533	.030	.028	.948
	80%	.094	.016	.015	.914	.376	.026	.025	.954	.530	.027	.026	.946
MRIC_{AIC}	60%	.090	.020	.017	.891	.375	.034	.030	.955	.536	.035	.031	.943
	70%	.092	.017	.016	.894	.375	.029	.027	.954	.533	.030	.028	.946
	80%	.094	.016	.015	.914	.376	.026	.025	.954	.530	.027	.026	.947
MRIC_{HL}	60%	.090	.021	.018	.894	.375	.036	.030	.964	.535	.037	.031	.952
	70%	.092	.018	.016	.908	.375	.030	.027	.960	.533	.031	.028	.950
	80%	.094	.016	.015	.917	.376	.026	.025	.954	.530	.027	.026	.943
MRNNMI	60%	.090	.020	.018	.888	.375	.035	.030	.960	.535	.036	.031	.949
	70%	.092	.018	.016	.900	.375	.030	.027	.957	.533	.031	.028	.947
	80%	.094	.016	.015	.918	.376	.026	.025	.954	.530	.027	.026	.948
DRNNMI	60%	.090	.019	.017	.876	.375	.033	.030	.958	.535	.034	.031	.943
	70%	.092	.017	.016	.900	.375	.029	.027	.948	.533	.030	.028	.940
	80%	.094	.016	.015	.913	.376	.026	.025	.952	.530	.027	.026	.941
MRPMMI	60%	.090	.028	.019	.917	.375	.048	.033	.980	.535	.050	.034	.975
	70%	.092	.023	.017	.921	.375	.039	.029	.971	.533	.040	.030	.967
	80%	.094	.019	.015	.932	.376	.031	.026	.974	.530	.032	.027	.965
$\text{MICE}_{\text{default}}$	60%	.106	.017	.024	.814	.374	.025	.031	.874	.519	.025	.030	.886
	70%	.104	.016	.020	.878	.375	.024	.027	.911	.521	.024	.027	.917
	80%	.102	.015	.017	.915	.376	.023	.024	.938	.522	.024	.025	.928
MICE_{CM}	60%	.106	.017	.024	.811	.375	.025	.031	.867	.520	.025	.031	.883
	70%	.103	.016	.020	.877	.375	.024	.027	.914	.522	.024	.027	.918
	80%	.102	.015	.017	.909	.376	.023	.024	.935	.522	.024	.025	.930
MICE_{INM}	60%	.103	.016	.023	.829	.368	.024	.031	.863	.528	.025	.030	.881
	70%	.102	.016	.019	.883	.371	.024	.027	.912	.528	.024	.027	.915
	80%	.101	.015	.017	.914	.373	.023	.024	.934	.526	.024	.025	.929
$\text{MICE}_{\text{INNM}}$	60%	.115	.018	.022	.823	.463	.026	.030	.112	.422	.025	.026	.027
	70%	.111	.017	.019	.875	.441	.025	.027	.279	.448	.024	.025	.134
	80%	.108	.016	.017	.917	.421	.024	.025	.547	.472	.024	.024	.418

Note. In Scenario 2 one of the outcome models is correctly defined and all the response models are incorrectly defined. *CCA* refers to the complete-case analysis, *RR* to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Similar to the results with $n = 50$, the bias in the response seems to be small, as the AEPs of the CCA are close to the true population proportions. As observed earlier, the robust methods seem to perform similarly in terms of AEPs for all the categories of Y. This indicates that weighting based on model quality did not reduce the bias of the estimates. In contrast to the previous findings with $n = 50$, the estimates of the MICE variants, except $\text{MICE}_{\text{INNM}}$, are closer to the true population proportions than the estimates of the robust methods, especially for categories 1 and 3. For the smallest category, robust methods seem to underestimate the number of units belonging to the category for all the response rates, whereas for the largest category the opposite occurs. The AEPs for all the MI methods seem to get closer to the true proportions as the response rate increases.

In line with previous findings, the ASEs of all the robust methods slightly overestimate their true SDs for all the categories, whereas the ASEs of the MICE variants underestimated their SDs. All the MICE variants have lower ASEs compared to the robust methods for all the response rates. Of all the MI methods MRPMMI results in the largest ASE. For all the methods, ASE decreases as the response rate increases. The robust methods perform similarly regarding SD, except MRPMMI has slightly higher SDs for the small response rates. For the smallest category, robust estimates resulted in smaller SDs compared to the MICE variants, whereas for the larger categories the SDs for all the methods were closer to each other.

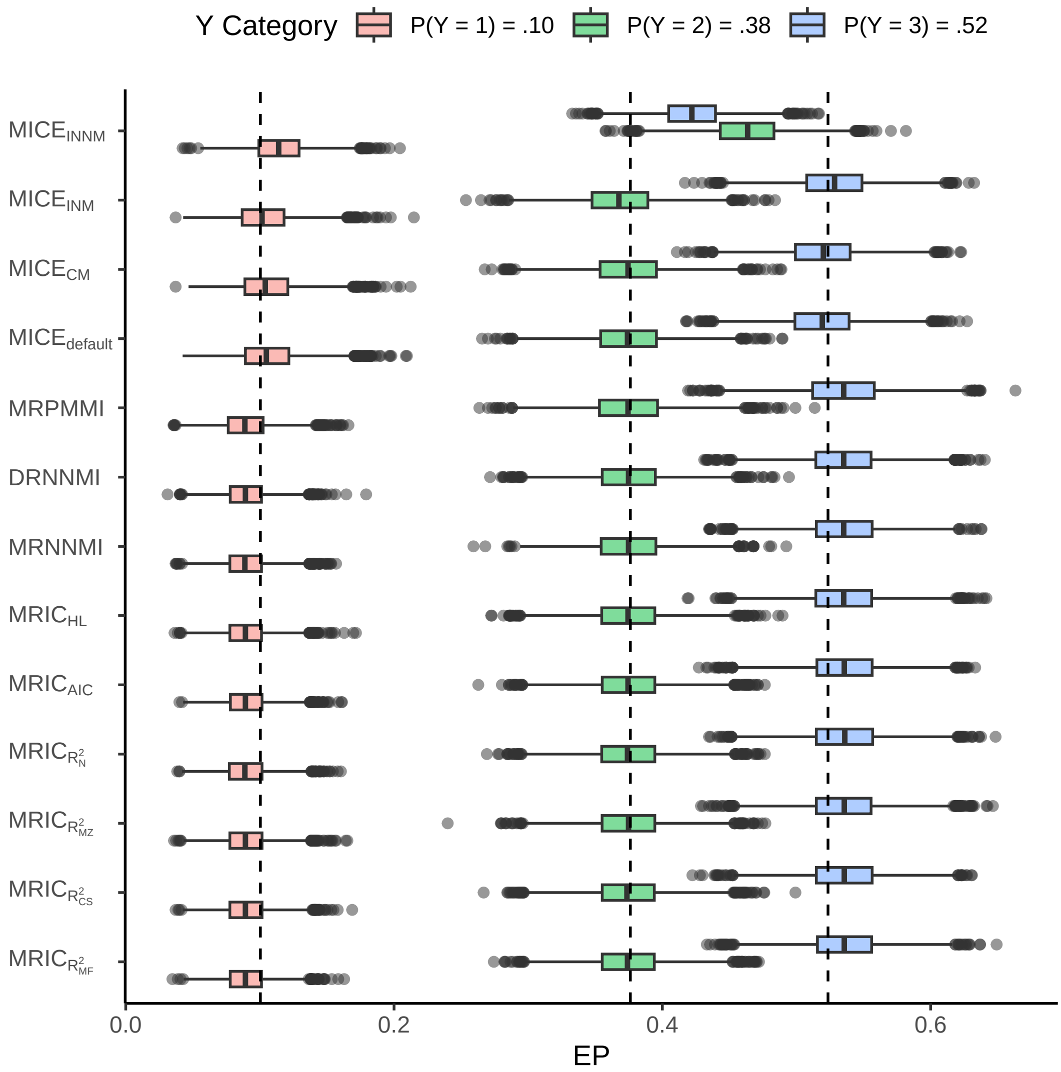
In terms of CR, most robust methods perform similarly, and seem to be able to better estimate larger categories as the CRs are closer to .950. The only exception is once again MRPMMI, which has a larger CR than other robust methods likely due to a larger ASE. The MICE variants that produce more unbiased estimates than the robust methods ($\text{MICE}_{\text{default}}$, MICE_{CM} and MICE_{INM}) have lower CRs than the robust methods, likely due to having lower ASEs. Of all the methods $\text{MICE}_{\text{INNM}}$ again has the worst performance, especially for the more prevalent categories, resulting in very small CRs and biased estimates.

To summarise, the proposed methods performed similarly as the existing robust methods, indicating that the weighting approach based on model quality did not reduce the bias of the estimates or improve accuracy. In contrast with $n = 50$, for $n = 500$ the MICE variants, apart from $\text{MICE}_{\text{INNM}}$, performed better in terms of bias compared to robust methods for all response rates. Similarly to earlier findings, $\text{MICE}_{\text{INNM}}$ resulted in the most biased estimates of all the methods. In terms of precision, the robust methods had lower SDs compared to the MICE variants for the smallest category, while the methods had similar SDs for the category 2, and the MICE variants had slightly lower SDs for the largest category. Of all the MI methods, MRPMMI resulted in the largest SDs for the larger categories. Similar to previous findings, the robust methods resulted in slightly overestimated ASEs, whereas the MICE variants underestimated them.

Figures 7-9 demonstrate the distribution of the estimated proportions for all the MI methods for each response rate.

Figure 7.

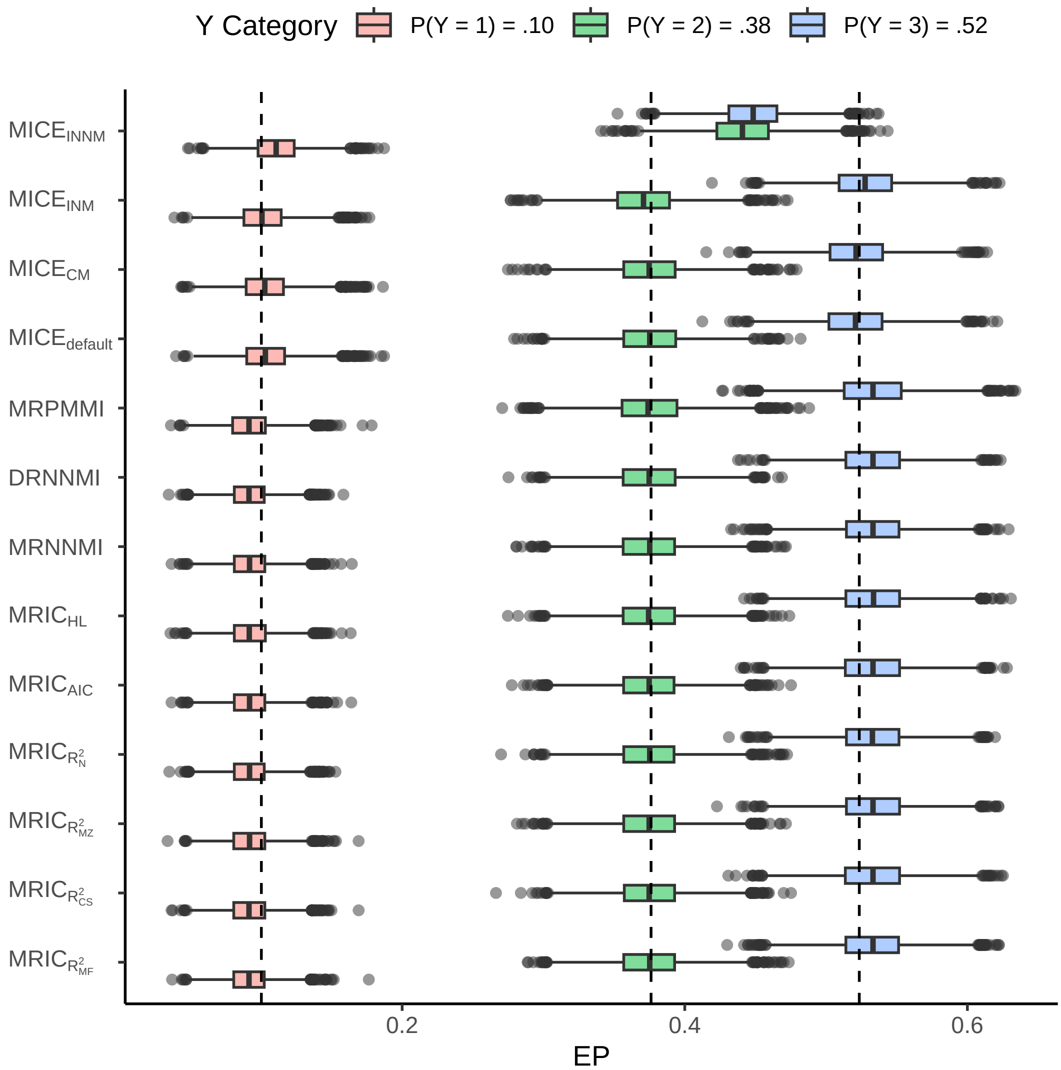
The Distributions of the Estimated Proportions for the 5000 Samples When $RR = 60\%$



Note. The figure shows the distributions of the estimated proportions in Scenario 2 for $n = 500$ when $RR = 60\%$. EP refers to the estimated proportion and RR to the response rate.

Figure 8.

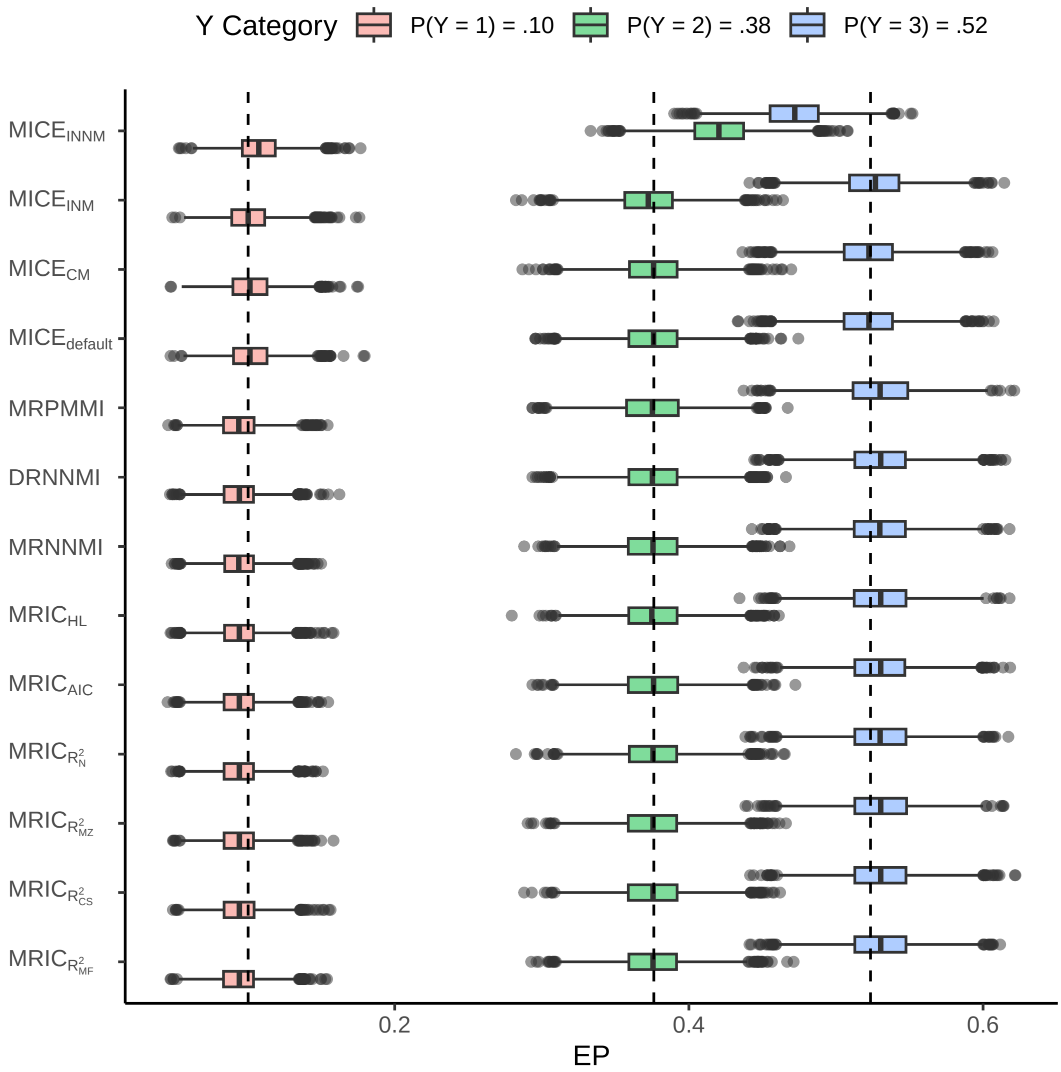
The Distributions of the Estimated Proportions for the 5000 Samples When $RR = 70\%$



Note. The figure shows the distributions of the estimated proportions in Scenario 2 for $n = 500$ when $RR = 70\%$. EP refers to the estimated proportion and RR to the response rate.

Figure 9.

The Distributions of the Estimated Proportions for the 5000 Samples When $RR = 80\%$



Note. The figure shows the distributions of the estimated proportions in scenario 2 for $n = 500$ when $RR = 80\%$. EP refers to the estimated proportion and RR to the response rate.

The figures show that when compared to $n = 50$ (see Figures 4-6), the distribution of the estimates for each MI method becomes narrower as the sample size increases. As earlier stated, it can be seen that the AEPs of the MICE variants, apart from $\text{MICE}_{\text{INNM}}$, are closer to the true values than the robust methods. As was observed with $n = 50$, $\text{MICE}_{\text{INNM}}$ again seems to have difficulty differentiating between the two largest categories as the distributions overlap noticeably.

Chapter 5

Discussion

The aim of this thesis was to examine if the proposed method, multiply robust imputation for categorical data (MRIC), performs better compared to the existing MI methods to impute missing categorical data in a simulation study. Whereas the existing robust MI methods applied prespecified weights to weight the predictive values, in MRIC the predictive values were weighted based on the working models' quality measures. Namely, three model quality measures were used: four types of pseudo-R^2, the Hosmer-Lemeshow chi-square test statistic and Akaike weights. However, contrary to the expectations, none of the weighting methods had a noticeable influence on the bias or precision of the estimates as all the robust methods performed similarly under all the conditions. This was especially unexpected for the conditions with small sample sizes and low response rates, as both pseudo-R^2 and the Hosmer-Lemeshow test are known to be sensitive to the small sample sizes and other aspects of the study design (Hemmert et al., 2018; Hosmer et al., 2013; Kramer & Zimmerman, 2007).

The results showed that all the robust methods produced estimates that were relatively close to the true population estimates even for smaller sample sizes and response rates. However, the robust methods only outperformed the MICE variants, except $\text{MICE}_{\text{INNM}}$, in terms of bias when the sample sizes were small. When the sample size was 200 or larger, the robust methods did not show any noticeable improvement in the accuracy of the estimates. In contrast, most of the estimates of the MICE variants showed improvement also for larger sample sizes. For sample sizes 500 and 1000 the estimates of the MICE variants, apart from $\text{MICE}_{\text{INNM}}$, were closer to the true population proportions in most of the conditions compared to the robust methods. In addition, the increase in the response rate seemed to improve the estimates of MICE variants more compared to the robust methods. Of all the MI methods, $\text{MICE}_{\text{INNM}}$ resulted in the most biased estimates in all the conditions.

In terms of the precision of the estimates, the results are less consistent. All the robust MI methods tended to overestimate the variance of the estimates whereas the MICE variants underestimated the variance, especially for lower response rates. The robust methods had relatively small differences in the precision of the estimates when observing the true SD, apart from MRPMMI, which resulted in slightly lower precision for the lower response rates. When the performance of the MI methods was evaluated with small and larger sample sizes, the results indicated that the estimates of the robust methods were more precise for the smallest category of the target variable compared to the MICE variants for both sample sizes. The estimates of MICE variants mostly had higher precision compared to the robust methods for the larger categories of the target variable. Generally, the precision of the estimates for all the methods improved as the response rate and sample size increased.

These results are partly in line with Breemer (2022) who found that MRNNMI and DRNNMI produced similar, slightly biased estimates for the categorical variable in a simulation study with a sample size of 300. In addition, it was reported that MICE, even with an incorrect outcome model, outperformed MRNNMI and DRNNMI in most conditions in terms of bias and average standard error. Similar results could also be observed in the current study for the MICE variants, apart from $\text{MICE}_{\text{INNM}}$, when the sample size was 500 or larger. In contrast, when the sample size was 200 or smaller, the results of the current study indicate that the MICE variants resulted in more biased estimates, and the MICE variants were also less precise for the smallest category compared to the robust methods in most conditions. The significantly worse performance of $\text{MICE}_{\text{INNM}}$ compared to the other MI methods can likely be attributed to the applied incorrect nonnested outcome model. In contrast, Breemer (2022) seemed to use incorrect nested outcome models more similar to this study's MICE_{INM}. In the current study, MICE_{INM} produced slightly more biased estimates while overall performing relatively similar to $\text{MICE}_{\text{default}}$ and MICE_{CM}.

The results of the thesis showed that there was no meaningful difference in the performance of the robust models in the different working model scenarios. This indicates that all the robust methods perform equally well regardless of whether the correct outcome and response models are defined or not. This is also in line with the findings by Breemer (2022) who reported that MRNNMI and DRNNMI performed similarly regardless of one or more incorrect working models being used. However, these findings are partly contradictory to the findings of other studies. Zhou et al. (2017) reported the increased bias of the estimates and reduction of precision in DRNNMI when an incorrect working model was specified. This is notable, as findings similar to Zhou et al. (2017) have also been reported in other studies that applied doubly robust MI approaches (Hsu et al., 2014; Long et al., 2012).

Overall, based on the results, weighting does not seem to have an influence on imputation performance for categorical data when multiple outcome and response models are defined, as MRIC performed similarly to MRNNMI. Furthermore, all the robust MI methods seem to perform similarly, apart from MRPMMI which results in slightly lower precision for larger sample sizes. Therefore, the best choice of MI method seems to be influenced more by the study design and available data. The results of the current study indicate that any of the robust MI methods would likely provide more stable estimates compared to MICE when the sample size is small and the response rate is low. In contrast, with larger sample sizes, MICE will likely result in less biased and more precise estimates. However, as earlier stated, MICE seems to be substantially influenced by the correct or incorrect nesting of the model. Without any prior knowledge of the correct working models, using MICE with the default settings without specifying a model is likely the more reliable option.

Several limitations of the current study have to be taken into consideration. First, the current study examined the performance of the MI methods only on one population from which the samples showed very small response bias. Therefore, the complete-case analysis resulted in very similar estimates as the robust methods. If one would use samples from a population in which more response bias was introduced, for example by increasing the correlation between the outcome variable and response indicator, it is possible that larger differences between the robust methods could be observed in different conditions. This might also be more in line with real world scenarios.

Second, a new sample was drawn every time the Hosmer-Lemeshow test was unable to divide the sample into at least two distinctive groups. As can be seen in Table D1 in Appendix D, a different number of samples was drawn between each working model scenario. This indicates that the samples between the working model scenarios slightly differ for sample sizes 200 or smaller. Although this did not seem to have influenced the main performance of the methods, as can be seen from the performance of the MICE variants between the different working model scenarios, this might have had some impact on the coverage rates that vary slightly for the same method between the working model scenarios.

Third, the study only examined the impact of using weighting based on working model quality in MRNNMI, but not in other preexisting robust MI methods. For all the other existing robust MI methods prespecified equal weights were used. Different weighting strategies have been shown to influence the performance of the doubly robust MI methods (Hsu et al., 2014; Long et al., 2012; Zhou et al., 2017), although there is no consensus of an optimal way to define the weights beyond the notion that they should be nonnegative. It is possible that if the weighting scenarios based on model quality measures were examined on other robust MI methods, for example using pseudo-R^2 with DRNNMI or Akaike weights with MRPMMI, larger differences between the robust methods could be observed in different working model scenarios.

Finally, the study did not examine many factors in the study design that are likely to influence the performance of the MI methods. As previously stated, with a sample size of 200 or larger, none of the robust method showed notable improvement in the estimates or precision compared to the results with small sample sizes. However, as the highest sample size in the current study was 1000, it is possible that improvement in the method would have been observed if a much higher sample size, for example 10000, would have been used. The study also did not examine the influence of a different number of predictors or the number of categories in the outcome variable; these are both known to influence some of the model quality measures. Furthermore, as was done in Breemer (2022), the current study only used working models with logit link functions. Considering this, the results cannot be generalised to models with different link functions.

In conclusion, this study has provided information about using weighting based on model quality measures in multiply robust MI for categorical data. Further studies are still needed to examine ways to implement weighting approaches based on the working model quality measures for other robust MI methods. As previously stated, weighting scenarios have been studied for doubly robust MI methods (Hsu et al., 2014; Long et al., 2012; Zhou et al., 2017). However, it would appear that weighting based on model quality measures has not been examined, at least not in the literature studied for this thesis. Future studies are also required to examine the implementation of other model quality measures for robust MI methods. For example, as the Hosmer-Lemeshow test does not perform reliably with very small sample sizes, other model quality measures, like the Pearson chi-square statistic could be an option if the data is in a correct format. Overall, more information about the impact of varying the number of explanatory variables, categories of the outcome variable and using models with different link functions on the robust MI methods with different weighting methods is needed. Finally, as the current study was limited to comparing the performance between the robust MI methods and MICE, future studies could compare the robust MI methods with different weighting approaches and other common methods to handle missing categorical data. For example, one other common imputation method from the class of donor methods is hot-deck imputation (Little & Rubin, 2020). Another option is to use a calibration estimator as was done by Zhou et al. (2017) or the multiply robust imputation method by Chen and Haziza (2017) that applies a calibration approach.

Bibliography

Akande, O., Li, F., & Reiter, J. (2017). An empirical comparison of multiple imputation methods for categorical data. *The American Statistician*, *71*(2), 162–170. https://doi.org/10.1080/00031305.2016.1277158

Anderson, D. R. (2008). *Model based inference in the life sciences: A primer on evidence*. Springer New York. https://doi.org/https://doi.org/10.1007/978-0-387-74075-1

Beullens, K., Loosveldt, G., Vandenplans, C., & Stoop, I. (2018). Response rates in the European Social Survey: Increasing, decreasing, or a matter of fieldwork efforts? *Survey Methods: Insights from the Field*, 1–12. https://doi.org/10.13094/SMIF-2018-00003

Breemer, L. (2022). *Multiply robust nearest neighbor multiple imputation for categorical missing data* (Master's thesis). Leiden University.

Chen, S., & Haziza, D. (2017). Multiply robust imputation procedures for the treatment of item nonresponse in surveys. *Biometrika*, *104*(2), 439–453. https://doi.org/10.1093/biomet/asx007

Chen, S., & Haziza, D. (2023). Doubly and multiply robust procedures for missing survey data. *The Survey Statistician*, *88*, 75–85.

Chen, S., Haziza, D., & Stubblefield, A. (2021). A note on multiply robust predictive mean matching imputation with complex survey data. *Survey Methodology*, *47*(1), 215–223. http://www.statcan.gc.ca/pub/12-001-x/2021001/article/00009-eng.htm

Cox, D. R., & Snell, E. J. (1989). *Flexible imputation of missing data.* (2nd ed.). Chapman and Hall.

de Leeuw, E., Hox, J., & Luiten, A. (2018). International nonresponse trends across countries and years: An analysis of 36 years of labour force survey data. *Survey Methods: Insights from the Field*, 1–11. https://doi.org/10.13094/SMIF-2018-00008

DeMaris, A. (2002). Explained variance in logistic regression: A Monte Carlo study of proposed measures. *Sociological Methods & Research*, *31*(1), 27–74. https://doi.org/10.1177/0049124102031001002

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Graham, J. W. (2012). *Missing data: Analysis and design*. Springer Science Business Media.

Hemmert, G. A. J., Schons, L. M., Wieseke, J., & Schimmelpfennig, H. (2018). Log-likelihood-based pseudo-R^2 in logistic regression: Deriving sample-sensitive benchmarks. *Sociological Methods & Research*, *47*(3), 507–531. https://doi.org/10.1177/0049124116638107

Hoetker, G. (2007). The use of logit and probit models in strategic management research: Critical issues. *Strategic Management Journal*, *28*(4), 331–343. https://doi.org/10.1002/smj.582

Hosmer, J., Lemeshow, D. W., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). John Wiley Sons.

Hsu, C.-H., Long, Q., Li, Y., & Jacobs, E. (2014). A nonparametric multiple imputation approach for data with missing covariate values with application to colorectal adenoma data.

Journal of Biopharmaceutical Statistics, *27*(3), 634–648. https://doi.org/10.1080/10543406.2014.888444
Kramer, A. A., & Zimmerman, J. E. (2007). Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. *Critical care medicine*, *35*(9), 2052–2056. https://doi.org/10.1097/01.CCM.0000275267.64078.B0
Lele, S. R., Keim, J. L., & Solymos, P. (2019). *ResourceSelection: Resource selection (probability) functions for use-availability data* [R package version 0.3-5]. https://CRAN.R-project.org/package=ResourceSelection
Little, R. J. A., & Rubin, D. B. (2020). *Statistical analysis with missing data* (3rd ed.). John Wiley & Sons.
Long, Q., Hsu, C.-H., & Li, Y. (2012). Doubly robust nonparametric multiple imputation for ignorable missing data. *Statistica Sinica*, *22*(3), 149–172.
McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). Academic Press.
McFadden, D. (1979). Quantitative methods for analysing travel behaviour of individuals: Some recent developments. In D. A. Hensher & P. R. Stopher (Eds.), *Behavioural travel modelling* (pp. 279–318). Croom Helm.
McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, *4*(1), 103–120. https://doi.org/10.1080/0022250X.1975.9989847
Menard, S. (2002). *Applied logistic regression analysis* (2nd ed.). Sage.
Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, *78*(3), 691–692. https://doi.org/10.1093/biomet/78.3.691
Paul, P., Pennell, M. L., & Lemeshow, S. (2013). Standardizing the power of the Hosmer–Lemeshow goodness of fit test in large data sets. *Statistics in medicine*, *32*(1), 67–80. https://doi.org/10.1002/sim.5525
Peytchev, A. (2012). Consequences of survey nonresponse. *The ANNALS of the American Academy of Political and Social Science*, *645*(1), 88–111. https://doi.org/10.1177/0002716212461748
R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/
Rubin, D. B. (1987a). The calculation of posterior distributions by data augmentation: Comment: A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The SIR algorithm. *Journal of the American Statistical Association*, *82*(398), 543–546. https://doi.org/10.2307/2289460
Rubin, D. B. (1987b). *Multiple imputation for nonresponse in surveys*. Wiley.
Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. https://doi.org/10.1037//1082-989X.7.2.147
Smith, T. J., & McKenna, C. M. (2013). A comparison of logistic regression pseudo R^2 indices. *Multiple Linear Regression Viewpoints*, *39*(2), 17–526.
Stavseth, M. R., Clausen, T., & Røislien, J. (2019). How handling missing data may impact conclusions: A comparison of six different imputation methods for categorical questionnaire data. *SAGE Open Medicine*, *7*, 1–12. https://doi.org/10.1177/2050312118822912
Van Buuren, S. (2018). *Flexible imputation of missing data*. CRC Press.
Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

van der Palm, D. W., van der Ark, L. A., & Vermunt, J. K. (2012). A comparison of incomplete-data methods for categorical data. *Statistical methods in medical research*, *25*(2), 754–774. https://doi.org/10.1177/0962280212465502

Veall, M. R., & Zimmermann, K. F. (1996). Pseudo-R^2 measures for some common limited dependent variable models. *Journal of Economic Surveys*, *10*(3), 241–259. https://doi.org/10.1111/j.1467-6419.1996.tb00013.x

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (Fourth) [ISBN 0-387-95457-0]. Springer. https://www.stats.ox.ac.uk/pub/MASS4/

Vermunt, J. K., van Ginkel, J. R., van der Ark, L. A., & Sijtsma, K. (2008). Multiple imputation of incomplete categorical data using latent class analysis. *Sociological Methodology*, *38*(1), 369–397. https://doi.org/10.1111/j.1467-9531.2008.00202.x

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A grammar of data manipulation* [R package version 1.1.2]. https://CRAN.R-project.org/package=dplyr

Windmeijer, F. A. (1995). Goodness-of-fit measures in binary choice models. *Econometric reviews*, *14*(1), 101–116. https://doi.org/10.1080/07474939508800306

Zhou, M., He, Y., Yu, M., & Hsu, C.-H. (2017). A nonparametric multiple imputation approach for missing categorical data. *BMC Medical Research Methodology*, *17*(1), 1–12. https://doi.org/10.1186/s12874-017-0360-2

Appendix A

The R scripts for the simulation study can be found on https://github.com/JonnaTeinonen/MastersThesis

Appendix B

Table B1.
The Outcome and Response Models Used for Each Robust Method in Scenario 1

		Scenario 1		
Method		Outcome models		Response Models
MRIC	CM	$Y \sim X_1 + X_2 + X_3$	CM	$R \sim X_1 + X_2 + X_3$
	INM	$Y \sim X_1 + X_3$	INM	$R \sim X_1 + X_2$
	INNM	$Y \sim X_1 + X_2 + X_4 + X_3 : X_4$	INNM	$R \sim X_1 + X_3 + X_4 + X_2 : X_4$
MRNNMI	CM	$Y \sim X_1 + X_2 + X_3$	CM	$R \sim X_1 + X_2 + X_3$
	INM	$Y \sim X_1 + X_3$	INM	$R \sim X_1 + X_2$
	INNM	$Y \sim X_1 + X_2 + X_4 + X_3 : X_4$	INNM	$R \sim X_1 + X_3 + X_4 + X_2 : X_4$
DRNNMI	CM	$Y \sim X_1 + X_2 + X_3$	CM	$R \sim X_1 + X_2 + X_3$
MRPMMI	CM	$Y \sim X_1 + X_2 + X_3$		-
	INM	$Y \sim X_1 + X_3$		-
	INNM	$Y \sim X_1 + X_2 + X_4 + X_3 : X_4$		-

Note. In Scenario 1 one of the outcome and response models is correctly defined. *CM* refers to the correct model, *INM* to an incorrect nested model and *INNM* to an incorrect nonnested model.

Table B2.

The Outcome and Response Models Used for Each Robust Method in Scenario 2

		Scenario 2		
Method		Outcome models		Response Models
MRIC	CM	$Y \sim X_1 + X_2 + X_3$	INM1	$R \sim X_1 + X_2$
	INM	$Y \sim X_1 + X_3$	INM2	$R \sim X_1 + X_3$
	INNM	$Y \sim X_1 + X_2 + X_4 + X_3 : X_4$	INNM	$R \sim X_1 + X_3 + X_4 + X_2 : X_4$
MRNNMI	CM	$Y \sim X_1 + X_2 + X_3$	INM1	$R \sim X_1 + X_2$
	INM	$Y \sim X_1 + X_3$	INM2	$R \sim X_1 + X_3$
	INNM	$Y \sim X_1 + X_2 + X_4 + X_3 : X_4$	INNM	$R \sim X_1 + X_3 + X_4 + X_2 : X_4$
DRNNMI	CM	$Y \sim X_1 + X_2 + X_3$	INM1	$R \sim X_1 + X_2$
MRPMMI	CM	$Y \sim X_1 + X_2 + X_3$		-
	INM	$Y \sim X_1 + X_3$		-
	INNM	$Y \sim X_1 + X_2 + X_4 + X_3 : X_4$		-

Note. In Scenario 2 one of the outcome models is correctly defined but all the response models are incorrect. *CM* refers to the correct model, *INM* to an incorrect nested model and *INNM* to an incorrect nonnested model.

Table B3.
The Outcome and Response Models Used for Each Robust Method in Scenario 3

Method		Scenario 3 Outcome models		Response Models
MRIC	INM1	$Y \sim X_1 + X_3$	CM	$R \sim X_1 + X_2 + X_3$
	INM2	$Y \sim X_1 + X_2$	INM	$R \sim X_1 + X_2$
	INNM	$Y \sim X_1 + X_2 + X_4 + X_3 : X_4$	INNM	$R \sim X_1 + X_3 + X_4 + X_2 : X_4$
MRNNMI	INM1	$Y \sim X_1 + X_3$	CM	$R \sim X_1 + X_2 + X_3$
	INM2	$Y \sim X_1 + X_2$	INM	$R \sim X_1 + X_2$
	INNM	$Y \sim X_1 + X_2 + X_4 + X_3 : X_4$	INNM	$R \sim X_1 + X_3 + X_4 + X_2 : X_4$
DRNNMI	INM1	$Y \sim X_1 + X_3$	CM	$R \sim X_1 + X_2 + X_3$
MRPMMI	INM1	$Y \sim X_1 + X_3$		-
	INM2	$Y \sim X_1 + X_2$		-
	INNM	$Y \sim X_1 + X_2 + X_4 + X_3 : X_4$		-

Note. In Scenario 3 one of the response models is correctly defined but all the outcome models are incorrect. *CM* refers to the correct model, *INM* to an incorrect nested model and *INNM* to an incorrect nonnested model.

Table B4.
The Outcome and Response Models Used for Each Robust Method in Scenario 4

Method		Scenario 4 Outcome models		Response Models
MRIC	INM1	$Y \sim X_1 + X_3$	INM1	$R \sim X_1 + X_2$
	INM2	$Y \sim X_1 + X_2$	INM2	$R \sim X_1 + X_3$
	INM3	$Y \sim X_2 + X_3$	INM3	$R \sim X_2 + X_3$
MRNNMI	INM1	$Y \sim X_1 + X_3$	INM1	$R \sim X_1 + X_2$
	INM2	$Y \sim X_1 + X_2$	INM2	$R \sim X_1 + X_3$
	INM3	$Y \sim X_2 + X_3$	INM3	$R \sim X_2 + X_3$
DRNNMI	INM1	$Y \sim X_1 + X_3$	INM1	$R \sim X_1 + X_2$
MRPMMI	INM1	$Y \sim X_1 + X_3$		-
	INM2	$Y \sim X_1 + X_2$		-
	INM3	$Y \sim X_2 + X_3$		-

Note. In Scenario 4 all the outcome and response models are incorrect nested models. *INM* refers to an incorrect nested model.

Table B5.
The Outcome and Response Models Used for Each Robust Method in Scenario 5

		Scenario 5		
Method		Outcome models		Response Models
MRIC	INNM1	$Y \sim X_1 + X_2 + X_4 + X_3 : X_4$	INNM1	$R \sim X_1 + X_3 + X_4 + X_2 : X_4$
	INNM2	$Y \sim X_1 + X_2 + X_4$	INNM2	$R \sim X_1 + X_3 + X_4$
	INNM3	$Y \sim X_1 + X_3 + X_1 : X_4$	INNM3	$R \sim X_2 + X_3 + X_3 : X_4$
MRNNMI	INNM1	$Y \sim X_1 + X_2 + X_4 + X_3 : X_4$	INNM1	$R \sim X_1 + X_3 + X_4 + X_2 : X_4$
	INNM2	$Y \sim X_1 + X_2 + X_4$	INNM2	$R \sim X_1 + X_3 + X_4$
	INNM3	$Y \sim X_1 + X_3 + X_1 : X_4$	INNM3	$R \sim X_2 + X_3 + X_3 : X_4$
DRNNMI	INNM1	$Y \sim X_1 + X_2 + X_4 + X_3 : X_4$	INNM1	$R \sim X_1 + X_3 + X_4 + X_2 : X_4$
MRPMMI	INNM1	$Y \sim X_1 + X_2 + X_4 + X_3 : X_4$		-
	INNM2	$Y \sim X_1 + X_2 + X_4$		-
	INNM3	$Y \sim X_1 + X_3 + X_1 : X_4$		-

Note. In Scenario 5 all the outcome and response models are incorrect nonnested models. *INMM* refers to an incorrect nonnested model.

Table B6.
The Outcome Models Used With Different MICE Variations

Method	Outcome models	Response Models
$\text{MICE}_{\text{default}}$	-	-
MICE_{CM}	$Y \sim X_1 + X_2 + X_3$	-
MICE_{INM}	$Y \sim X_1 + X_3$	-
$\text{MICE}_{\text{INNM}}$	$Y \sim X_1 + X_2 + X_4 + X_3 : X_4$	-

Note. The four different MICE variants use the same models in all conditions. *CM* refers to the correct model, *INM* to an incorrect nested model and *INNM* to an incorrect nonnested model.

Appendix C

Influence of the Predictors on the Outcome Variable

Table C1.
The Impact of Each Predictor on the Outcome Variable Y

	$n = 50$				$n = 1000$			
Predictors	p-value	SD	ASR	LRT	p-value	SD	ASR	LRT
X_1	.028 *	.08	.88	13.82	$<$ .001 *	$<$ .01	1.00	208.85
X_2	.008 *	.02	.96	17.38	$<$ .001 *	$<$ .01	1.00	277.00
X_3	.017 *	.07	.93	15.25	$<$ .001 *	$<$ .01	1.00	250.98
X_4	.394	.29	.13	3.13	.484	.29	.06	2.17

Note. The results indicate the average p-value, average standard deviation (SD) of the p-values, average significance rate (ASR) of the p-values with $\alpha = .05$ and the average Likelihood Ratio Test (LRT) statistics for each predictor over 300 drawn samples by using the type III LRT.

Influence of the Predictors on the Response Indicator

Table C2.
The Impact of Each Predictor on the Response Indicator When $n = 50$

	$ARR = .60$				$ARR = .70$				$ARR = .79$			
Predictors	p-value	SD	ASR	LRT	p-value	SD	ASR	LRT	p-value	SD	ASR	LRT
X_1	.193	.24	.39	3.87	.235	.28	.38	3.92	.265	.32	.35	3.61
X_2	$<$ .001 *	$<$ .01	1.00	24.01	.001 *	.02	.99	21.57	.005 *	.03	.98	18.44
X_3	.001 *	$<$ .01	1.00	21.98	.004 *	.03	.99	20.02	.010 *	.08	.98	17.12
X_4	.482	.32	.11	1.35	.440	.30	.09	1.47	.504	.32	.10	1.80

Note. ARR refers to the average response rate of the 300 samples. The results indicate the average p-value, average standard deviation (SD) of the p-values, average significance rate (ASR) of the p-values with $\alpha = .05$ and the average Likelihood Ratio Test (LRT) statistics for each predictor over 300 drawn samples by using the type III LRT.

Table C3.
The Impact of Each Predictor on the Response Indicator When n = 100

	ARR = .60				*ARR* = .70				*ARR* = .79			
Predictors	p-value	SD	ASR	LRT	p-value	SD	ASR	LRT	p-value	SD	ASR	LRT
X_1	.090	.17	.63	6.14	.129	.22	.59	5.87	.163	.24	.47	4.96
X_2	$<$ .001 *	$<$ .01	1.00	44.83	$<$ .001 *	$<$ .01	1.00	40.75	$<$ .001 *	$<$ .01	1.00	34.96
X_3	$<$ .001 *	$<$ .01	1.00	41.51	$<$ .001 *	$<$ .01	1.00	37.64	$<$ .001 *	$<$ .01	1.00	32.11
X_4	.472	.30	.08	1.24	.442	.29	.09	1.36	.450	.30	.10	1.35

Note. ARR refers to the average response rate of the 300 samples. The results indicate the average p-value, average standard deviation (SD) of the p-values, average significance rate (ASR) of the p-values with $\alpha = .05$ and the average Likelihood Ratio Test (LRT) statistics for each predictor over 300 drawn samples by using the type III LRT.

Table C4.
The Impact of Each Predictor on the Response Indicator When n = 200

	ARR = .60				*ARR* = .70				*ARR* = .79			
Predictors	p-value	SD	ASR	LRT	p-value	SD	ASR	LRT	p-value	SD	ASR	LRT
X_1	.022 *	.08	.91	11.20	.026 *	.09	.91	11.01	.053	.13	.80	9.04
X_2	$<$ .001 *	$<$ .01	1.00	87.56	$<$ .001 *	$<$ .01	1.00	78.78	$<$ .001 *	$<$ .01	1.00	65.10
X_3	$<$ .001 *	$<$ .01	1.00	82.52	$<$ .001 *	$<$ .01	1.00	76.25	$<$ .001 *	$<$ .01	1.00	62.91
X_4	.473	.29	.05	1.09	.487	.30	.07	1.12	.486	.29	.06	1.09

Note. ARR refers to the average response rate of the 300 samples. The results indicate the average p-value, average standard deviation (SD) of the p-values, average significance rate (ASR) of the p-values with $\alpha = .05$ and the average Likelihood Ratio Test (LRT) statistics for each predictor over 300 drawn samples by using the type III LRT.

Table C5.
The Impact of Each Predictor on the Response Indicator When n = 500

	ARR = .60				*ARR* = .70				*ARR* = .80			
Predictors	p-value	SD	ASR	LRT	p-value	SD	ASR	LRT	p-value	SD	ASR	LRT
X_1	$<$ .001 *	$<$ .01	1.00	27.57	$<$ .001 *	$<$ .01	1.00	25.20	.002 *	.01	.99	20.14
X_2	$<$ .001 *	$<$ .01	1.00	218.25	$<$ .001 *	$<$ .01	1.00	196.47	$<$ .001 *	$<$ .01	1.00	163.03
X_3	$<$ .001 *	$<$ .01	1.00	203.77	$<$ .001 *	$<$ .01	1.00	183.85	$<$ .001 *	$<$ .01	1.00	152.27
X_4	.510	.28	.05	.95	.496	.28	.03	.95	.530	.28	.05	.83

Note. ARR refers to the average response rate of the 300 samples. The results indicate the average p-value, average standard deviation (SD) of the p-values, average significance rate (ASR) of the p-values with $\alpha = .05$ and the average Likelihood Ratio Test (LRT) statistics for each predictor over 300 drawn samples by using the type III LRT.

Table C6.

The Impact of Each Predictor on the Response Indicator When n = 1000

	ARR = .60				*ARR* = .70				*ARR* = .80			
Predictors	p-value	SD	ASR	LRT	p-value	SD	ASR	LRT	p-value	SD	ASR	LRT
X_1	< .001 *	< .01	1.00	54.22	< .001 *	< .01	1.00	48.97	< .001 *	< .01	1.00	40.04
X_2	< .001 *	< .01	1.00	437.76	< .001 *	< .01	1.00	395.22	< .001 *	< .01	1.00	324.98
X_3	< .001 *	< .01	1.00	407.52	< .001 *	< .01	1.00	367.43	< .001 *	< .01	1.00	306.04
X_4	.475	.29	.06	1.07	.524	.03	.04	0.94	.479	.29	.06	1.10

Note. ARR refers to the average response rate of the 300 samples. The results indicate the average p-value, average standard deviation (SD) of the p-values, average significance rate (ASR) of the p-values with $\alpha = .05$ and the average Likelihood Ratio Test (LRT) statistics for each predictor over 300 drawn samples by using the type III LRT.

Appendix D

Table D1.
The Hosmer-Lemeshow Error Check and Average Response Rate for Each Sample

		n = 50		n = 100		n = 200		n = 500		n = 1000	
Model Scenario	RR	HL error	ARR	HL error	ARR	HL error	ARR	HL error	ARR	HL error	ARR
Scenario 1	60%	.007	.604	.011	.603	.002	.603	0	.603	0	.603
	70%	.012	.705	.007	.704	.001	.705	0	.705	0	.704
	80%	.014	.799	.005	.798	.001	.799	0	.799	0	.799
Scenario 2	60%	.007	.604	.011	.603	.002	.603	0	.603	0	.603
	70%	.012	.705	.007	.704	.001	.705	0	.705	0	.704
	80%	.014	.799	.005	.798	.001	.799	0	.799	0	.799
Scenario 3	60%	.007	.604	.010	.603	.003	.603	0	.603	0	.603
	70%	.009	.705	.009	.704	.001	.705	0	.705	0	.704
	80%	.013	.798	.008	.798	.001	.799	0	.799	0	.799
Scenario 4	60%	.015	.604	.043	.603	.016	.603	0	.603	0	.603
	70%	.019	.705	.040	.704	.011	.705	0	.705	0	.704
	80%	.032	.798	.034	.798	.003	.799	0	.799	0	.799
Scenario 5	60%	.005	.604	.007	.603	.001	.603	0	.603	0	.603
	70%	.008	.705	.007	.704	0	.705	0	.705	0	.704
	80%	.013	.798	.005	.798	0	.799	0	.799	0	.799

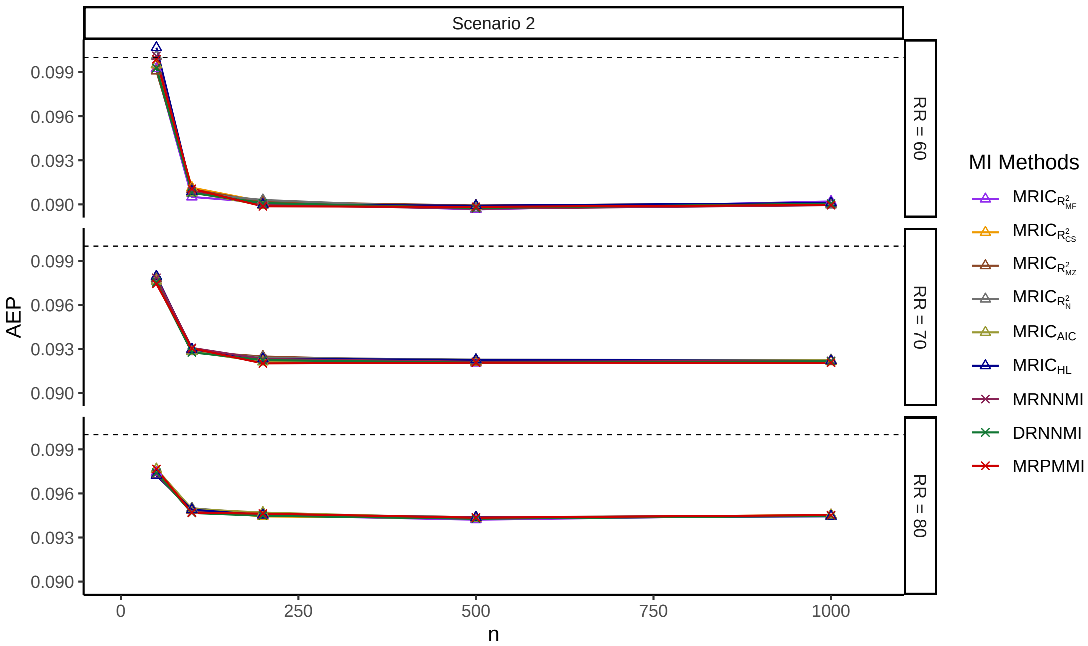
Note. Scenarios refer to different working model scenarios that were introduced in section 4.2. *HL error* denotes the fraction of how often a new sample had to be drawn during the 5000 iterations because the Hosmer-Lemeshow test was unable to divide the units in the sample into at least two different groups. *ARR* denotes the average response rate for the final selected 5000 samples.

Appendix E

AEP for Robust Methods in Scenario 2

Figure E1.

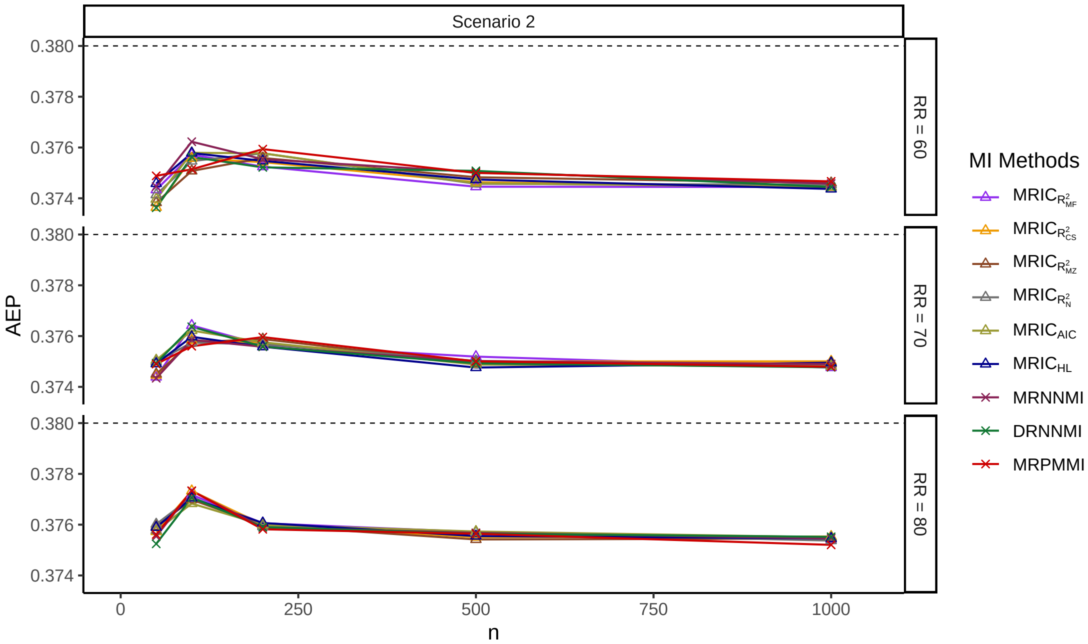
The Average Estimated Proportion for Category 1 of Y for the Robust Methods in Scenario 2



Note. AEP refers to the average estimated proportion for category 1 and *RR* to the response rate. The y-axis scale for AEP has been adjusted to illustrate how all the robust methods had a similar AEP.

Figure E2.

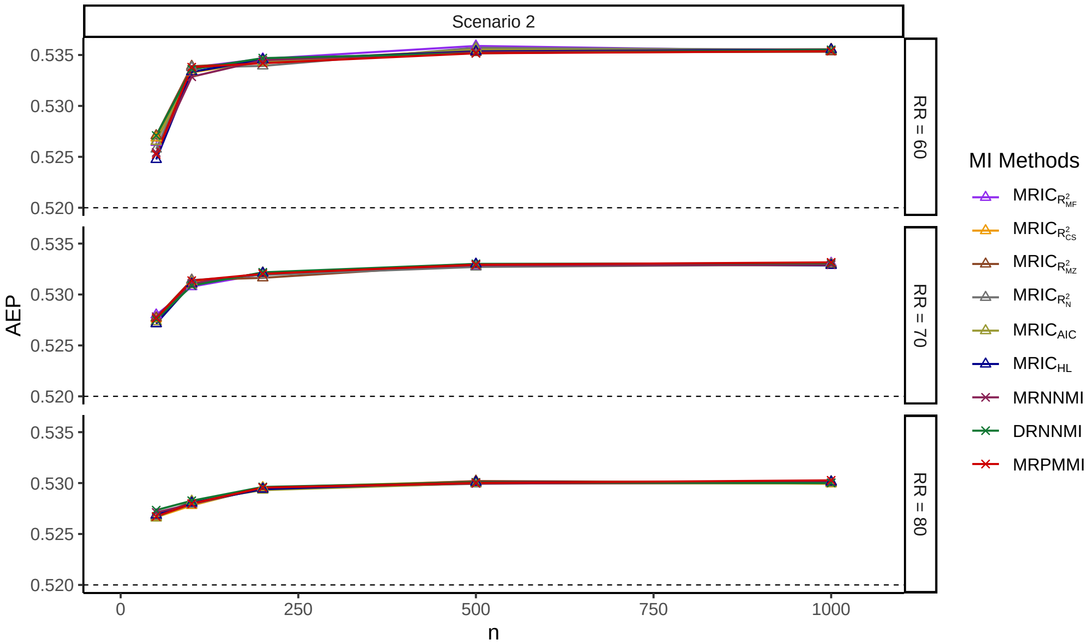
The Average Estimated Proportion for Category 2 of Y for the Robust Methods in Scenario 2



Note. *AEP* refers to the average estimated proportion for category 2 and *RR* to the response rate. The y-axis scale for AEP has been adjusted to illustrate how all the robust methods had a similar AEP.

Figure E3.

The Average Estimated Proportion for Category 3 of Y for the Robust Methods in Scenario 2



Note. *AEP* refers to the average estimated proportion for category 3 and *RR* to the response rate. The y-axis scale for AEP has been adjusted to illustrate how all the robust methods had a similar AEP.

Appendix F

Tables for Scenario 1 for all the MI Methods

Table F1.

The Results of the Different Robust MI Methods in Scenario 1 for $n = 50$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.096	-	.049	-	.375	-	.088	-	.529	-	.091	-
	70%	.096	-	.047	-	.375	-	.082	-	.529	-	.085	-
	80%	.096	-	.045	-	.376	-	.076	-	.528	-	.079	-
MRIC$_{R^2_{MF}}$	60%	.099	.060	.050	.923	.374	.100	.093	.942	.527	.104	.095	.947
	70%	.098	.053	.048	.913	.374	.089	.085	.938	.528	.092	.088	.940
	80%	.097	.048	.046	.910	.376	.080	.078	.942	.527	.083	.082	.943
MRIC$_{R^2_{CS}}$	60%	.100	.060	.051	.925	.373	.100	.092	.939	.527	.103	.095	.943
	70%	.098	.053	.048	.917	.374	.088	.085	.939	.528	.091	.088	.947
	80%	.097	.048	.046	.912	.376	.080	.078	.943	.527	.083	.081	.942
MRIC$_{R^2_{MZ}}$	60%	.099	.060	.050	.926	.374	.100	.093	.938	.527	.104	.096	.941
	70%	.098	.053	.048	.915	.375	.089	.085	.937	.527	.092	.088	.940
	80%	.097	.048	.046	.911	.376	.080	.077	.940	.527	.083	.081	.943
MRIC$_{R^2_{N}}$	60%	.100	.061	.051	.932	.374	.101	.092	.940	.526	.104	.095	.943
	70%	.097	.053	.048	.916	.375	.089	.086	.938	.528	.092	.089	.940
	80%	.097	.048	.046	.910	.376	.080	.078	.942	.527	.083	.081	.944
MRIC$_{\text{AIC}}$	60%	.099	.059	.051	.923	.374	.098	.092	.937	.526	.101	.095	.941
	70%	.098	.052	.048	.911	.375	.087	.085	.936	.527	.090	.088	.937
	80%	.098	.048	.046	.911	.376	.079	.078	.940	.527	.082	.081	.939
MRIC$_{\text{HL}}$	60%	.101	.068	.053	.919	.375	.112	.095	.942	.525	.116	.098	.949
	70%	.098	.058	.050	.908	.375	.097	.087	.946	.527	.100	.090	.948
	80%	.097	.050	.046	.903	.376	.085	.079	.948	.527	.087	.082	.952
MRNNMI	60%	.100	.061	.051	.928	.375	.100	.092	.939	.525	.103	.096	.943
	70%	.098	.053	.048	.917	.374	.089	.085	.942	.528	.092	.088	.942
	80%	.097	.048	.046	.913	.375	.080	.078	.942	.527	.083	.081	.942
DRNNMI	60%	.100	.059	.051	.924	.374	.098	.092	.934	.526	.101	.095	.939
	70%	.098	.052	.048	.914	.375	.087	.085	.939	.528	.090	.088	.934
	80%	.097	.048	.046	.909	.376	.080	.077	.946	.527	.082	.081	.941
MRPMMI	60%	.100	.061	.051	.928	.375	.102	.093	.938	.525	.105	.096	.939
	70%	.097	.053	.048	.914	.375	.090	.085	.943	.528	.093	.088	.943
	80%	.098	.048	.046	.912	.376	.081	.078	.942	.527	.083	.081	.943
MICE$_{\text{default}}$	60%	.163	.060	.076	.814	.362	.074	.081	.905	.475	.076	.079	.886
	70%	.136	.054	.063	.895	.369	.073	.077	.925	.495	.075	.076	.923
	80%	.120	.050	.053	.927	.373	.072	.073	.937	.507	.074	.075	.936
MICE$_{\text{CM}}$	60%	.154	.058	.076	.837	.364	.075	.083	.904	.483	.077	.082	.886
	70%	.131	.053	.062	.897	.369	.073	.078	.917	.500	.075	.078	.917
	80%	.117	.049	.053	.925	.373	.072	.074	.933	.510	.074	.076	.937
MICE$_{\text{INM}}$	60%	.151	.058	.067	.875	.358	.074	.078	.913	.491	.077	.081	.908
	70%	.129	.053	.058	.925	.365	.073	.075	.927	.506	.076	.078	.929
	80%	.116	.049	.051	.930	.370	.072	.073	.930	.514	.074	.075	.935
MICE$_{\text{INNM}}$	60%	.138	.057	.067	.888	.446	.079	.089	.818	.416	.077	.079	.686
	70%	.124	.053	.059	.915	.432	.077	.083	.873	.444	.075	.078	.791
	80%	.115	.049	.053	.920	.417	.075	.077	.909	.468	.074	.075	.872

Note. In Scenario 1 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Table F2.

The Results of the Different Robust MI Methods in Scenario 1 for $n = 100$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.090	-	.036	-	.376	-	.063	-	.534	-	.066	-
	70%	.093	-	.035	-	.376	-	.057	-	.531	-	.060	-
	80%	.095	-	.033	-	.377	-	.054	-	.528	-	.056	-
$\text{MRIC}_{R^2_{MF}}$	60%	.091	.043	.039	.883	.376	.075	.067	.949	.534	.077	.070	.950
	70%	.093	.038	.036	.891	.376	.065	.060	.954	.531	.067	.063	.947
	80%	.095	.034	.033	.901	.377	.058	.056	.949	.528	.059	.058	.948
$\text{MRIC}_{R^2_{CS}}$	60%	.091	.043	.039	.884	.376	.075	.067	.946	.533	.077	.070	.947
	70%	.093	.038	.036	.894	.376	.065	.060	.949	.531	.067	.062	.953
	80%	.095	.034	.033	.905	.377	.058	.055	.948	.528	.059	.058	.948
$\text{MRIC}_{R^2_{MZ}}$	60%	.091	.043	.039	.883	.375	.075	.068	.950	.534	.077	.070	.944
	70%	.093	.038	.036	.895	.376	.065	.060	.949	.531	.067	.063	.945
	80%	.095	.034	.034	.901	.377	.058	.056	.948	.528	.059	.058	.946
$\text{MRIC}_{R^2_{N}}$	60%	.091	.043	.039	.879	.375	.075	.067	.950	.534	.077	.071	.942
	70%	.093	.038	.036	.898	.376	.065	.060	.949	.531	.067	.062	.952
	80%	.095	.034	.033	.897	.377	.058	.056	.951	.528	.060	.058	.950
MRIC_{AIC}	60%	.091	.042	.038	.883	.376	.073	.066	.948	.534	.075	.069	.944
	70%	.093	.037	.036	.893	.376	.064	.060	.952	.531	.066	.062	.949
	80%	.095	.034	.033	.905	.377	.057	.055	.948	.528	.059	.057	.950
MRIC_{HL}	60%	.091	.049	.040	.887	.376	.087	.070	.960	.533	.089	.073	.965
	70%	.093	.041	.037	.895	.376	.070	.062	.961	.531	.073	.064	.958
	80%	.095	.036	.034	.903	.377	.060	.056	.959	.528	.062	.058	.957
MRNNMI	60%	.091	.043	.039	.885	.376	.075	.068	.943	.533	.077	.070	.943
	70%	.093	.038	.036	.892	.376	.065	.060	.949	.531	.067	.063	.949
	80%	.095	.034	.034	.899	.377	.058	.055	.955	.528	.060	.057	.952
DRNNMI	60%	.091	.042	.039	.881	.376	.073	.067	.947	.533	.075	.070	.948
	70%	.093	.037	.036	.893	.376	.064	.060	.951	.531	.066	.062	.945
	80%	.095	.034	.033	.904	.377	.057	.055	.951	.528	.059	.057	.948
MRPMMI	60%	.091	.045	.039	.888	.375	.078	.068	.951	.534	.081	.070	.955
	70%	.093	.039	.037	.898	.376	.067	.060	.952	.531	.069	.063	.951
	80%	.095	.035	.034	.904	.377	.059	.056	.950	.528	.061	.058	.951
$\text{MICE}_{\text{default}}$	60%	.133	.040	.057	.808	.369	.054	.065	.885	.498	.055	.064	.882
	70%	.118	.037	.045	.884	.374	.053	.059	.911	.508	.054	.059	.914
	80%	.110	.034	.038	.913	.376	.051	.054	.928	.514	.053	.055	.925
MICE_{CM}	60%	.127	.039	.056	.814	.370	.054	.067	.877	.502	.055	.065	.876
	70%	.115	.036	.045	.879	.374	.053	.059	.911	.511	.054	.060	.912
	80%	.108	.034	.038	.911	.376	.051	.055	.932	.516	.053	.055	.928
MICE_{INM}	60%	.126	.039	.052	.848	.364	.054	.063	.886	.510	.055	.064	.892
	70%	.114	.036	.043	.893	.369	.053	.057	.912	.517	.054	.059	.924
	80%	.108	.034	.037	.914	.373	.051	.054	.934	.520	.053	.055	.930
$\text{MICE}_{\text{INNM}}$	60%	.122	.039	.050	.863	.457	.057	.066	.676	.420	.055	.059	.524
	70%	.115	.037	.043	.896	.439	.055	.059	.779	.446	.054	.056	.679
	80%	.111	.035	.038	.913	.420	.053	.056	.858	.469	.053	.054	.809

Note. In Scenario 1 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Table F3.
The Results of the Different Robust MI Methods in Scenario 1 for $n = 200$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.090	-	.026	-	.375	-	.045	-	.535	-	.046	-
	70%	.092	-	.024	-	.376	-	.041	-	.532	-	.042	-
	80%	.095	-	.023	-	.376	-	.038	-	.529	-	.040	-
$\text{MRIC}_{R^2_{MF}}$	60%	.090	.032	.028	.895	.375	.054	.048	.948	.534	.056	.050	.948
	70%	.092	.028	.025	.911	.375	.047	.043	.956	.532	.048	.044	.952
	80%	.095	.025	.024	.914	.376	.041	.039	.955	.529	.043	.041	.947
$\text{MRIC}_{R^2_{CS}}$	60%	.090	.032	.028	.902	.375	.055	.048	.955	.534	.056	.049	.947
	70%	.092	.028	.026	.905	.376	.047	.043	.951	.532	.048	.044	.951
	80%	.094	.025	.024	.915	.376	.041	.040	.949	.529	.043	.041	.946
$\text{MRIC}_{R^2_{MZ}}$	60%	.090	.032	.028	.894	.376	.055	.048	.954	.535	.057	.049	.948
	70%	.092	.028	.026	.905	.376	.047	.043	.958	.532	.048	.044	.955
	80%	.095	.025	.024	.915	.376	.041	.040	.951	.529	.042	.041	.948
$\text{MRIC}_{R^2_{N}}$	60%	.090	.032	.028	.896	.375	.055	.048	.952	.534	.056	.050	.953
	70%	.092	.028	.026	.903	.376	.047	.043	.956	.532	.048	.044	.952
	80%	.095	.025	.024	.915	.376	.041	.040	.947	.529	.042	.041	.950
MRIC_{AIC}	60%	.090	.031	.028	.889	.376	.053	.048	.951	.534	.055	.049	.946
	70%	.092	.027	.025	.909	.376	.046	.043	.955	.532	.047	.044	.948
	80%	.095	.024	.024	.910	.376	.041	.039	.950	.529	.042	.041	.945
MRIC_{HL}	60%	.090	.034	.029	.903	.376	.059	.049	.957	.534	.061	.050	.962
	70%	.092	.028	.026	.911	.376	.049	.043	.958	.532	.050	.044	.958
	80%	.094	.025	.024	.918	.376	.042	.040	.949	.529	.043	.041	.949
MRNNMI	60%	.090	.032	.028	.897	.375	.055	.048	.951	.535	.057	.049	.946
	70%	.092	.028	.026	.911	.375	.047	.043	.954	.532	.048	.044	.955
	80%	.094	.025	.024	.913	.376	.041	.040	.948	.530	.043	.041	.947
DRNNMI	60%	.090	.031	.028	.891	.376	.053	.048	.948	.534	.054	.049	.945
	70%	.092	.027	.026	.904	.376	.046	.042	.953	.532	.047	.044	.947
	80%	.095	.024	.024	.910	.376	.041	.039	.951	.530	.042	.041	.949
MRPMMI	60%	.090	.035	.029	.904	.376	.061	.050	.962	.534	.063	.050	.964
	70%	.092	.030	.026	.910	.376	.052	.044	.965	.532	.053	.045	.963
	80%	.095	.026	.024	.920	.376	.044	.040	.960	.530	.045	.041	.956
$\text{MICE}_{\text{default}}$	60%	.116	.027	.039	.819	.373	.039	.048	.875	.511	.039	.047	.881
	70%	.109	.025	.031	.884	.375	.038	.042	.911	.516	.038	.042	.919
	80%	.105	.024	.027	.912	.375	.037	.039	.936	.520	.037	.039	.933
MICE_{CM}	60%	.114	.027	.039	.820	.373	.039	.048	.876	.513	.039	.048	.885
	70%	.108	.025	.031	.878	.375	.038	.043	.913	.517	.038	.042	.918
	80%	.104	.024	.027	.909	.375	.037	.039	.935	.520	.037	.039	.934
MICE_{INM}	60%	.113	.026	.037	.833	.366	.038	.047	.868	.521	.039	.047	.895
	70%	.106	.025	.030	.887	.370	.037	.042	.913	.523	.038	.042	.924
	80%	.103	.024	.026	.913	.372	.036	.038	.937	.524	.037	.039	.935
$\text{MICE}_{\text{INNM}}$	60%	.117	.028	.035	.852	.461	.041	.047	.445	.421	.039	.042	.271
	70%	.113	.026	.030	.900	.440	.039	.042	.621	.447	.038	.039	.485
	80%	.109	.025	.027	.915	.420	.038	.039	.777	.471	.037	.038	.708

Note. In Scenario 1 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Table F4.

The Results of the Different Robust MI Methods in Scenario 1 for $n = 500$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.090	-	.016	-	.375	-	.027	-	.535	-	.029	-
	70%	.092	-	.015	-	.375	-	.026	-	.533	-	.027	-
	80%	.094	-	.014	-	.376	-	.024	-	.530	-	.025	-
$\text{MRIC}_{R^2_{MF}}$	60%	.090	.021	.018	.888	.375	.036	.030	.961	.536	.037	.031	.950
	70%	.092	.018	.016	.906	.375	.030	.027	.957	.533	.031	.028	.952
	80%	.094	.016	.015	.916	.376	.026	.025	.959	.530	.027	.026	.949
$\text{MRIC}_{R^2_{CS}}$	60%	.090	.021	.018	.894	.374	.036	.030	.963	.536	.037	.032	.949
	70%	.092	.018	.016	.905	.375	.030	.027	.957	.533	.031	.028	.949
	80%	.094	.016	.015	.916	.376	.026	.025	.959	.530	.027	.026	.948
$\text{MRIC}_{R^2_{MZ}}$	60%	.090	.021	.018	.896	.375	.036	.030	.959	.535	.037	.031	.950
	70%	.092	.018	.016	.906	.375	.030	.027	.959	.533	.031	.028	.950
	80%	.094	.016	.015	.920	.375	.026	.025	.956	.530	.027	.026	.946
$\text{MRIC}_{R^2_{N}}$	60%	.090	.021	.018	.892	.375	.036	.030	.966	.536	.037	.031	.947
	70%	.092	.018	.016	.905	.375	.030	.027	.962	.533	.031	.028	.946
	80%	.094	.016	.015	.914	.376	.026	.025	.952	.530	.027	.026	.945
MRIC_{AIC}	60%	.090	.020	.018	.885	.375	.035	.030	.963	.535	.036	.031	.945
	70%	.092	.018	.016	.901	.375	.030	.027	.958	.533	.031	.028	.944
	80%	.094	.016	.015	.908	.376	.026	.025	.952	.530	.027	.026	.943
MRIC_{HL}	60%	.090	.022	.018	.896	.375	.038	.030	.966	.535	.039	.031	.956
	70%	.092	.019	.016	.908	.375	.031	.027	.958	.533	.032	.028	.952
	80%	.094	.016	.015	.915	.375	.027	.025	.959	.530	.028	.026	.948
MRNNMI	60%	.090	.021	.018	.892	.375	.036	.031	.958	.535	.038	.032	.953
	70%	.092	.018	.016	.899	.375	.031	.027	.964	.533	.032	.028	.950
	80%	.094	.016	.015	.916	.376	.026	.025	.957	.530	.027	.026	.948
DRNNMI	60%	.090	.020	.018	.885	.375	.034	.030	.957	.535	.035	.031	.944
	70%	.092	.018	.016	.900	.375	.030	.027	.960	.533	.030	.028	.946
	80%	.094	.016	.015	.919	.376	.026	.025	.958	.530	.027	.026	.947
MRPMMI	60%	.090	.028	.019	.917	.375	.048	.033	.980	.535	.050	.034	.975
	70%	.092	.023	.017	.921	.375	.039	.029	.971	.533	.040	.030	.967
	80%	.094	.019	.015	.932	.376	.031	.026	.974	.530	.032	.027	.965
$\text{MICE}_{\text{default}}$	60%	.106	.017	.024	.814	.374	.025	.031	.874	.519	.025	.030	.886
	70%	.104	.016	.020	.878	.375	.024	.027	.911	.521	.024	.027	.917
	80%	.102	.015	.017	.915	.376	.023	.024	.938	.522	.024	.025	.928
MICE_{CM}	60%	.106	.017	.024	.811	.375	.025	.031	.867	.520	.025	.031	.883
	70%	.103	.016	.020	.877	.375	.024	.027	.914	.522	.024	.027	.918
	80%	.102	.015	.017	.909	.376	.023	.024	.935	.522	.024	.025	.930
MICE_{INM}	60%	.103	.016	.023	.829	.368	.024	.031	.863	.528	.025	.030	.881
	70%	.102	.016	.019	.883	.371	.024	.027	.912	.528	.024	.027	.915
	80%	.101	.015	.017	.914	.373	.023	.024	.934	.526	.024	.025	.929
$\text{MICE}_{\text{INNM}}$	60%	.115	.018	.022	.823	.463	.026	.030	.112	.422	.025	.026	.027
	70%	.111	.017	.019	.875	.441	.025	.027	.279	.448	.024	.025	.134
	80%	.108	.016	.017	.917	.421	.024	.025	.547	.472	.024	.024	.418

Note. In Scenario 1 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Table F5.

The Results of the Different Robust MI Methods in Scenario 1 for $n = 1000$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.090	-	.012	-	.374	-	.020	-	.536	-	.020	-
	70%	.092	-	.011	-	.375	-	.018	-	.533	-	.019	-
	80%	.095	-	.010	-	.375	-	.017	-	.530	-	.017	-
$\text{MRIC}_{R^2_{MF}}$	60%	.090	.015	.013	.878	.374	.026	.022	.964	.535	.027	.022	.946
	70%	.092	.013	.012	.886	.375	.022	.019	.961	.533	.023	.020	.942
	80%	.095	.011	.011	.907	.375	.019	.018	.961	.530	.019	.018	.948
$\text{MRIC}_{R^2_{CS}}$	60%	.090	.015	.013	.872	.374	.026	.022	.963	.536	.027	.022	.944
	70%	.092	.013	.012	.882	.375	.022	.019	.961	.533	.022	.020	.943
	80%	.095	.011	.011	.907	.375	.019	.018	.962	.530	.019	.018	.948
$\text{MRIC}_{R^2_{MZ}}$	60%	.090	.015	.013	.879	.374	.026	.022	.967	.536	.027	.023	.942
	70%	.092	.013	.012	.889	.375	.022	.019	.965	.533	.023	.020	.943
	80%	.095	.011	.011	.904	.375	.019	.017	.962	.530	.019	.018	.949
$\text{MRIC}_{R^2_{N}}$	60%	.090	.015	.013	.871	.375	.026	.022	.960	.535	.027	.022	.940
	70%	.092	.013	.012	.884	.375	.022	.019	.965	.533	.023	.020	.942
	80%	.095	.011	.011	.906	.376	.019	.017	.960	.530	.019	.018	.947
MRIC_{AIC}	60%	.090	.015	.013	.857	.374	.025	.022	.960	.536	.026	.022	.938
	70%	.092	.013	.012	.875	.375	.021	.019	.966	.533	.022	.020	.937
	80%	.095	.011	.011	.901	.375	.019	.017	.965	.530	.019	.018	.945
MRIC_{HL}	60%	.090	.017	.013	.881	.374	.028	.022	.969	.536	.029	.023	.956
	70%	.092	.014	.012	.892	.375	.023	.019	.971	.533	.024	.020	.951
	80%	.094	.012	.011	.910	.375	.020	.018	.966	.530	.020	.018	.955
MRNNMI	60%	.090	.016	.013	.877	.375	.027	.022	.962	.535	.027	.023	.944
	70%	.092	.013	.012	.889	.375	.022	.019	.960	.533	.023	.020	.937
	80%	.095	.011	.011	.905	.375	.019	.018	.963	.530	.020	.018	.948
DRNNMI	60%	.090	.014	.013	.860	.374	.025	.021	.960	.536	.026	.022	.928
	70%	.092	.013	.012	.875	.375	.021	.019	.957	.533	.022	.020	.939
	80%	.094	.011	.011	.900	.375	.019	.017	.960	.530	.019	.018	.948
MRPMMI	60%	.090	.024	.015	.920	.375	.042	.026	.982	.535	.043	.027	.976
	70%	.092	.019	.013	.919	.375	.033	.022	.982	.533	.034	.023	.973
	80%	.095	.015	.011	.928	.375	.025	.019	.978	.530	.026	.019	.975
$\text{MICE}_{\text{default}}$	60%	.104	.012	.017	.812	.375	.017	.022	.871	.521	.018	.022	.880
	70%	.102	.011	.014	.876	.375	.017	.019	.912	.523	.017	.019	.915
	80%	.101	.011	.012	.922	.376	.016	.017	.939	.523	.017	.017	.941
MICE_{CM}	60%	.103	.012	.017	.811	.375	.017	.022	.873	.522	.018	.022	.880
	70%	.102	.011	.014	.872	.375	.017	.019	.911	.523	.017	.019	.918
	80%	.101	.011	.012	.922	.376	.016	.017	.937	.523	.017	.017	.943
MICE_{INM}	60%	.101	.011	.016	.825	.369	.017	.022	.853	.530	.018	.022	.862
	70%	.100	.011	.014	.870	.371	.017	.019	.903	.529	.017	.019	.907
	80%	.100	.011	.012	.918	.373	.016	.017	.934	.527	.017	.017	.941
$\text{MICE}_{\text{INNM}}$	60%	.114	.012	.016	.750	.464	.018	.021	.009	.422	.018	.019	.000
	70%	.111	.012	.014	.827	.441	.018	.019	.058	.448	.017	.018	.010
	80%	.108	.011	.012	.893	.421	.017	.018	.265	.471	.017	.017	.132

Note. In Scenario 1 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Tables for Scenario 2 for all the MI Methods

Table F6.

The Results of the Different Robust MI Methods in Scenario 2 for $n = 50$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.096	-	.049	-	.375	-	.088	-	.529	-	.091	-
	70%	.096	-	.047	-	.375	-	.082	-	.529	-	.085	-
	80%	.096	-	.045	-	.376	-	.076	-	.528	-	.079	-
$\text{MRIC}_{R^2_{MF}}$	60%	.099	.060	.051	.923	.374	.100	.093	.941	.526	.103	.095	.941
	70%	.098	.053	.048	.908	.374	.088	.085	.940	.528	.091	.088	.941
	80%	.097	.048	.046	.912	.376	.080	.078	.941	.527	.083	.081	.942
$\text{MRIC}_{R^2_{CS}}$	60%	.100	.060	.051	.927	.374	.099	.092	.934	.527	.102	.096	.942
	70%	.098	.053	.048	.915	.374	.088	.085	.940	.528	.091	.088	.942
	80%	.098	.048	.045	.915	.376	.080	.078	.941	.527	.082	.081	.943
$\text{MRIC}_{R^2_{MZ}}$	60%	.099	.060	.050	.926	.374	.100	.092	.940	.527	.103	.095	.940
	70%	.098	.053	.048	.916	.375	.088	.085	.939	.528	.091	.088	.942
	80%	.097	.048	.046	.908	.376	.080	.078	.942	.527	.083	.081	.941
$\text{MRIC}_{R^2_{N}}$	60%	.100	.060	.051	.922	.374	.100	.092	.940	.526	.104	.096	.941
	70%	.098	.053	.048	.914	.375	.089	.085	.940	.527	.091	.089	.941
	80%	.097	.048	.046	.909	.376	.080	.078	.947	.527	.083	.081	.944
MRIC_{AIC}	60%	.100	.058	.051	.923	.374	.097	.092	.936	.527	.101	.095	.944
	70%	.098	.052	.048	.914	.375	.087	.085	.938	.527	.090	.088	.936
	80%	.098	.048	.046	.913	.376	.079	.078	.941	.527	.082	.081	.939
MRIC_{HL}	60%	.101	.068	.053	.920	.375	.112	.095	.941	.525	.116	.098	.947
	70%	.098	.058	.050	.908	.375	.097	.087	.946	.527	.100	.090	.947
	80%	.097	.050	.046	.903	.376	.084	.079	.948	.527	.087	.082	.950
MRNNMI	60%	.100	.060	.051	.930	.374	.100	.093	.937	.525	.103	.096	.943
	70%	.098	.053	.048	.916	.374	.088	.085	.938	.528	.091	.088	.940
	80%	.097	.048	.046	.911	.376	.080	.078	.944	.527	.083	.081	.942
DRNNMI	60%	.099	.058	.050	.927	.374	.096	.091	.933	.527	.099	.094	.935
	70%	.098	.051	.048	.913	.375	.086	.084	.937	.528	.089	.087	.940
	80%	.097	.047	.046	.909	.375	.079	.078	.940	.527	.081	.081	.942
MRPMMI	60%	.100	.061	.051	.928	.375	.102	.093	.938	.525	.105	.096	.939
	70%	.097	.053	.048	.914	.375	.090	.085	.943	.528	.093	.088	.943
	80%	.098	.048	.046	.912	.376	.081	.078	.942	.527	.083	.081	.943
$\text{MICE}_{\text{default}}$	60%	.163	.060	.076	.814	.362	.074	.081	.905	.475	.076	.079	.886
	70%	.136	.054	.063	.895	.369	.073	.077	.925	.495	.075	.076	.923
	80%	.120	.050	.053	.927	.373	.072	.073	.937	.507	.074	.075	.936
MICE_{CM}	60%	.154	.058	.076	.837	.364	.075	.083	.904	.483	.077	.082	.886
	70%	.131	.053	.062	.897	.369	.073	.078	.917	.500	.075	.078	.917
	80%	.117	.049	.053	.925	.373	.072	.074	.933	.510	.074	.076	.937
MICE_{INM}	60%	.151	.058	.067	.875	.358	.074	.078	.913	.491	.077	.081	.908
	70%	.129	.053	.058	.925	.365	.073	.075	.927	.506	.076	.078	.929
	80%	.116	.049	.051	.930	.370	.072	.073	.930	.514	.074	.075	.935
$\text{MICE}_{\text{INNM}}$	60%	.138	.057	.067	.888	.446	.079	.089	.818	.416	.077	.079	.686
	70%	.124	.053	.059	.915	.432	.077	.083	.873	.444	.075	.078	.791
	80%	.115	.049	.053	.920	.417	.075	.077	.909	.468	.074	.075	.872

Note. In Scenario 2 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Table F7.
The Results of the Different Robust MI Methods in Scenario 2 for $n = 100$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.090	-	.036	-	.376	-	.063	-	.534	-	.066	-
	70%	.093	-	.035	-	.376	-	.057	-	.531	-	.060	-
	80%	.095	-	.033	-	.377	-	.054	-	.528	-	.056	-
$MRIC_{R^2_{MF}}$	60%	.091	.042	.038	.878	.376	.074	.067	.946	.534	.076	.070	.947
	70%	.093	.038	.036	.894	.376	.065	.060	.952	.531	.067	.062	.953
	80%	.095	.034	.033	.909	.377	.058	.055	.947	.528	.059	.058	.948
$MRIC_{R^2_{CS}}$	60%	.091	.043	.039	.887	.376	.074	.067	.949	.533	.076	.069	.947
	70%	.093	.038	.036	.895	.376	.064	.060	.952	.531	.066	.062	.946
	80%	.095	.034	.034	.900	.377	.058	.056	.946	.528	.059	.058	.943
$MRIC_{R^2_{MZ}}$	60%	.091	.043	.039	.880	.375	.074	.068	.946	.534	.077	.070	.946
	70%	.093	.038	.036	.891	.376	.065	.060	.947	.531	.067	.063	.948
	80%	.095	.034	.033	.904	.377	.057	.055	.945	.528	.059	.058	.947
$MRIC_{R^2_{N}}$	60%	.091	.042	.039	.883	.375	.074	.067	.950	.534	.077	.070	.949
	70%	.093	.038	.036	.896	.376	.064	.060	.949	.531	.066	.063	.946
	80%	.095	.034	.034	.903	.377	.058	.055	.951	.528	.059	.057	.952
$MRIC_{AIC}$	60%	.091	.041	.038	.884	.376	.072	.066	.947	.533	.074	.069	.943
	70%	.093	.037	.036	.891	.376	.063	.060	.950	.531	.065	.062	.950
	80%	.095	.034	.034	.899	.377	.057	.055	.945	.528	.059	.057	.947
$MRIC_{HL}$	60%	.091	.049	.040	.886	.376	.086	.070	.961	.533	.089	.073	.962
	70%	.093	.041	.037	.895	.376	.070	.062	.960	.531	.073	.064	.958
	80%	.095	.036	.034	.903	.377	.060	.056	.961	.528	.062	.058	.960
MRNNMI	60%	.091	.043	.039	.891	.376	.074	.068	.944	.533	.077	.070	.946
	70%	.093	.038	.036	.897	.376	.064	.060	.949	.531	.066	.062	.946
	80%	.095	.034	.034	.897	.377	.058	.055	.953	.528	.059	.057	.950
DRNNMI	60%	.091	.041	.038	.886	.376	.070	.066	.944	.534	.072	.069	.942
	70%	.093	.037	.036	.892	.376	.062	.060	.945	.531	.064	.062	.940
	80%	.095	.034	.033	.904	.377	.057	.055	.950	.528	.058	.057	.948
MRPMMI	60%	.091	.045	.039	.888	.375	.078	.068	.951	.534	.081	.070	.955
	70%	.093	.039	.037	.898	.376	.067	.060	.952	.531	.069	.063	.951
	80%	.095	.035	.034	.904	.377	.059	.056	.950	.528	.061	.058	.951
$MICE_{default}$	60%	.133	.040	.057	.808	.369	.054	.065	.885	.498	.055	.064	.882
	70%	.118	.037	.045	.884	.374	.053	.059	.911	.508	.054	.059	.914
	80%	.110	.034	.038	.913	.376	.051	.054	.928	.514	.053	.055	.925
$MICE_{CM}$	60%	.127	.039	.056	.814	.370	.054	.067	.877	.502	.055	.065	.876
	70%	.115	.036	.045	.879	.374	.053	.059	.911	.511	.054	.060	.912
	80%	.108	.034	.038	.911	.376	.051	.055	.932	.516	.053	.055	.928
$MICE_{INM}$	60%	.126	.039	.052	.848	.364	.054	.063	.886	.510	.055	.064	.892
	70%	.114	.036	.043	.893	.369	.053	.057	.912	.517	.054	.059	.924
	80%	.108	.034	.037	.914	.373	.051	.054	.934	.520	.053	.055	.930
$MICE_{INNM}$	60%	.122	.039	.050	.863	.457	.057	.066	.676	.420	.055	.059	.524
	70%	.115	.037	.043	.896	.439	.055	.059	.779	.446	.054	.056	.679
	80%	.111	.035	.038	.913	.420	.053	.056	.858	.469	.053	.054	.809

Note. In Scenario 2 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Table F8.

The Results of the Different Robust MI Methods in Scenario 2 for $n = 200$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.090	-	.026	-	.375	-	.045	-	.535	-	.046	-
	70%	.092	-	.024	-	.376	-	.041	-	.532	-	.042	-
	80%	.095	-	.023	-	.376	-	.038	-	.529	-	.040	-
$\text{MRIC}_{R^2_{MF}}$	60%	.090	.031	.028	.893	.375	.053	.048	.948	.535	.055	.050	.943
	70%	.092	.027	.025	.905	.376	.046	.043	.956	.532	.047	.044	.949
	80%	.095	.024	.024	.911	.376	.041	.039	.951	.529	.042	.041	.947
$\text{MRIC}_{R^2_{CS}}$	60%	.090	.031	.028	.895	.375	.053	.048	.947	.534	.055	.049	.949
	70%	.092	.027	.025	.905	.376	.046	.043	.952	.532	.048	.044	.948
	80%	.094	.024	.024	.909	.376	.041	.040	.949	.530	.042	.041	.948
$\text{MRIC}_{R^2_{MZ}}$	60%	.090	.031	.028	.894	.376	.054	.048	.951	.534	.056	.049	.949
	70%	.092	.027	.026	.900	.376	.046	.042	.956	.532	.048	.044	.952
	80%	.095	.024	.024	.913	.376	.041	.039	.949	.529	.042	.041	.949
$\text{MRIC}_{R^2_{N}}$	60%	.090	.031	.028	.895	.376	.054	.048	.951	.534	.056	.049	.951
	70%	.092	.027	.025	.908	.376	.046	.043	.955	.532	.048	.044	.954
	80%	.095	.024	.024	.917	.376	.041	.040	.948	.529	.042	.041	.945
MRIC_{AIC}	60%	.090	.030	.028	.888	.376	.052	.048	.950	.534	.053	.049	.946
	70%	.092	.027	.025	.900	.376	.045	.042	.955	.532	.047	.044	.949
	80%	.095	.024	.024	.911	.376	.041	.039	.949	.529	.042	.041	.948
MRIC_{HL}	60%	.090	.033	.028	.899	.375	.058	.049	.955	.535	.060	.050	.960
	70%	.092	.028	.026	.908	.376	.048	.043	.956	.532	.050	.044	.956
	80%	.095	.025	.024	.913	.376	.042	.040	.952	.529	.043	.041	.949
MRNNMI	60%	.090	.031	.028	.893	.376	.054	.048	.954	.534	.056	.049	.951
	70%	.092	.027	.025	.908	.376	.046	.043	.951	.532	.048	.044	.949
	80%	.095	.025	.024	.916	.376	.041	.040	.948	.530	.042	.041	.946
DRNNMI	60%	.090	.030	.027	.887	.375	.051	.047	.944	.535	.053	.048	.939
	70%	.092	.026	.025	.899	.376	.045	.043	.950	.532	.046	.044	.946
	80%	.094	.024	.024	.907	.376	.040	.040	.946	.530	.041	.041	.945
MRPMMI	60%	.090	.035	.029	.904	.376	.061	.050	.962	.534	.063	.050	.964
	70%	.092	.030	.026	.910	.376	.052	.044	.965	.532	.053	.045	.963
	80%	.095	.026	.024	.920	.376	.044	.040	.960	.530	.045	.041	.956
$\text{MICE}_{\text{default}}$	60%	.116	.027	.039	.819	.373	.039	.048	.875	.511	.039	.047	.881
	70%	.109	.025	.031	.884	.375	.038	.042	.911	.516	.038	.042	.919
	80%	.105	.024	.027	.912	.375	.037	.039	.936	.520	.037	.039	.933
MICE_{CM}	60%	.114	.027	.039	.820	.373	.039	.048	.876	.513	.039	.048	.885
	70%	.108	.025	.031	.878	.375	.038	.043	.913	.517	.038	.042	.918
	80%	.104	.024	.027	.909	.375	.037	.039	.935	.520	.037	.039	.934
MICE_{INM}	60%	.113	.026	.037	.833	.366	.038	.047	.868	.521	.039	.047	.895
	70%	.106	.025	.030	.887	.370	.037	.042	.913	.523	.038	.042	.924
	80%	.103	.024	.026	.913	.372	.036	.038	.937	.524	.037	.039	.935
$\text{MICE}_{\text{INNM}}$	60%	.117	.028	.035	.852	.461	.041	.047	.445	.421	.039	.042	.271
	70%	.113	.026	.030	.900	.440	.039	.042	.621	.447	.038	.039	.485
	80%	.109	.025	.027	.915	.420	.038	.039	.777	.471	.037	.038	.708

Note. In Scenario 2 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Table F9.

The Results of the Different Robust MI Methods in Scenario 2 for $n = 500$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.090	-	.016	-	.375	-	.027	-	.535	-	.029	-
	70%	.092	-	.015	-	.375	-	.026	-	.533	-	.027	-
	80%	.094	-	.014	-	.376	-	.024	-	.530	-	.025	-
$\text{MRIC}_{R^2_{MF}}$	60%	.090	.020	.017	.890	.374	.034	.029	.961	.536	.036	.030	.946
	70%	.092	.018	.016	.897	.375	.029	.027	.957	.533	.030	.028	.948
	80%	.094	.016	.015	.914	.376	.026	.025	.958	.530	.027	.026	.946
$\text{MRIC}_{R^2_{CS}}$	60%	.090	.020	.018	.890	.375	.034	.029	.959	.536	.036	.031	.947
	70%	.092	.017	.016	.905	.375	.029	.027	.951	.533	.030	.028	.942
	80%	.094	.016	.015	.915	.376	.026	.025	.954	.530	.027	.026	.946
$\text{MRIC}_{R^2_{MZ}}$	60%	.090	.020	.018	.892	.375	.035	.030	.954	.535	.036	.031	.943
	70%	.092	.017	.016	.901	.375	.029	.027	.956	.533	.030	.028	.945
	80%	.094	.016	.015	.915	.375	.026	.025	.952	.530	.027	.026	.943
$\text{MRIC}_{R^2_{N}}$	60%	.090	.020	.018	.883	.375	.034	.030	.957	.536	.036	.031	.944
	70%	.092	.018	.016	.900	.375	.029	.027	.957	.533	.030	.028	.948
	80%	.094	.016	.015	.914	.376	.026	.025	.954	.530	.027	.026	.946
MRIC_{AIC}	60%	.090	.020	.017	.891	.375	.034	.030	.955	.536	.035	.031	.943
	70%	.092	.017	.016	.894	.375	.029	.027	.954	.533	.030	.028	.946
	80%	.094	.016	.015	.914	.376	.026	.025	.954	.530	.027	.026	.947
MRIC_{HL}	60%	.090	.021	.018	.894	.375	.036	.030	.964	.535	.037	.031	.952
	70%	.092	.018	.016	.908	.375	.030	.027	.960	.533	.031	.028	.950
	80%	.094	.016	.015	.917	.376	.026	.025	.954	.530	.027	.026	.943
MRNNMI	60%	.090	.020	.018	.888	.375	.035	.030	.960	.535	.036	.031	.949
	70%	.092	.018	.016	.900	.375	.030	.027	.957	.533	.031	.028	.947
	80%	.094	.016	.015	.918	.376	.026	.025	.954	.530	.027	.026	.948
DRNNMI	60%	.090	.019	.017	.876	.375	.033	.030	.958	.535	.034	.031	.943
	70%	.092	.017	.016	.900	.375	.029	.027	.948	.533	.030	.028	.940
	80%	.094	.016	.015	.913	.376	.026	.025	.952	.530	.027	.026	.941
MRPMMI	60%	.090	.028	.019	.917	.375	.048	.033	.980	.535	.050	.034	.975
	70%	.092	.023	.017	.921	.375	.039	.029	.971	.533	.040	.030	.967
	80%	.094	.019	.015	.932	.376	.031	.026	.974	.530	.032	.027	.965
$\text{MICE}_{\text{default}}$	60%	.106	.017	.024	.814	.374	.025	.031	.874	.519	.025	.030	.886
	70%	.104	.016	.020	.878	.375	.024	.027	.911	.521	.024	.027	.917
	80%	.102	.015	.017	.915	.376	.023	.024	.938	.522	.024	.025	.928
MICE_{CM}	60%	.106	.017	.024	.811	.375	.025	.031	.867	.520	.025	.031	.883
	70%	.103	.016	.020	.877	.375	.024	.027	.914	.522	.024	.027	.918
	80%	.102	.015	.017	.909	.376	.023	.024	.935	.522	.024	.025	.930
MICE_{INM}	60%	.103	.016	.023	.829	.368	.024	.031	.863	.528	.025	.030	.881
	70%	.102	.016	.019	.883	.371	.024	.027	.912	.528	.024	.027	.915
	80%	.101	.015	.017	.914	.373	.023	.024	.934	.526	.024	.025	.929
$\text{MICE}_{\text{INNM}}$	60%	.115	.018	.022	.823	.463	.026	.030	.112	.422	.025	.026	.027
	70%	.111	.017	.019	.875	.441	.025	.027	.279	.448	.024	.025	.134
	80%	.108	.016	.017	.917	.421	.024	.025	.547	.472	.024	.024	.418

Note. In Scenario 2 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Table F10.
The Results of the Different Robust MI Methods in Scenario 2 for $n = 1000$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.090	-	.012	-	.374	-	.020	-	.536	-	.020	-
	70%	.092	-	.011	-	.375	-	.018	-	.533	-	.019	-
	80%	.095	-	.010	-	.375	-	.017	-	.530	-	.017	-
$\text{MRIC}_{R^2_{MF}}$	60%	.090	.015	.013	.866	.374	.025	.021	.957	.535	.025	.022	.933
	70%	.092	.012	.012	.875	.375	.021	.019	.959	.533	.022	.020	.943
	80%	.095	.011	.010	.908	.375	.018	.017	.959	.530	.019	.018	.948
$\text{MRIC}_{R^2_{CS}}$	60%	.090	.015	.013	.860	.374	.025	.021	.959	.536	.026	.022	.937
	70%	.092	.012	.012	.868	.375	.021	.019	.960	.533	.022	.020	.942
	80%	.095	.011	.011	.903	.376	.018	.017	.957	.530	.019	.018	.945
$\text{MRIC}_{R^2_{MZ}}$	60%	.090	.015	.013	.863	.375	.025	.021	.954	.535	.025	.022	.933
	70%	.092	.012	.012	.878	.375	.021	.019	.959	.533	.022	.020	.938
	80%	.094	.011	.011	.898	.375	.018	.017	.957	.530	.019	.018	.948
$\text{MRIC}_{R^2_N}$	60%	.090	.014	.013	.854	.375	.025	.021	.960	.536	.025	.022	.935
	70%	.092	.012	.012	.879	.375	.021	.019	.957	.533	.022	.020	.933
	80%	.094	.011	.011	.899	.375	.018	.017	.959	.530	.019	.018	.943
MRIC_{AIC}	60%	.090	.014	.013	.855	.374	.024	.021	.956	.535	.025	.022	.934
	70%	.092	.012	.012	.875	.375	.021	.019	.959	.533	.022	.020	.937
	80%	.095	.011	.011	.904	.376	.018	.017	.960	.530	.019	.018	.949
MRIC_{HL}	60%	.090	.015	.013	.868	.374	.026	.021	.966	.536	.026	.022	.946
	70%	.092	.013	.012	.885	.375	.022	.019	.960	.533	.022	.020	.945
	80%	.094	.011	.011	.907	.375	.019	.017	.963	.530	.019	.018	.945
MRNNMI	60%	.090	.015	.013	.866	.375	.025	.021	.963	.535	.026	.022	.939
	70%	.092	.013	.012	.880	.375	.021	.019	.957	.533	.022	.020	.937
	80%	.094	.011	.011	.907	.375	.019	.018	.958	.530	.019	.018	.945
DRNNMI	60%	.090	.014	.013	.850	.374	.024	.021	.955	.536	.025	.022	.930
	70%	.092	.012	.012	.873	.375	.021	.019	.959	.533	.021	.020	.937
	80%	.094	.011	.011	.899	.376	.018	.017	.962	.530	.019	.018	.947
MRPMMI	60%	.090	.024	.015	.920	.375	.042	.026	.982	.535	.043	.027	.976
	70%	.092	.019	.013	.919	.375	.033	.022	.982	.533	.034	.023	.973
	80%	.095	.015	.011	.928	.375	.025	.019	.978	.530	.026	.019	.975
$\text{MICE}_{\text{default}}$	60%	.104	.012	.017	.812	.375	.017	.022	.871	.521	.018	.022	.880
	70%	.102	.011	.014	.876	.375	.017	.019	.912	.523	.017	.019	.915
	80%	.101	.011	.012	.922	.376	.016	.017	.939	.523	.017	.017	.941
MICE_{CM}	60%	.103	.012	.017	.811	.375	.017	.022	.873	.522	.018	.022	.880
	70%	.102	.011	.014	.872	.375	.017	.019	.911	.523	.017	.019	.918
	80%	.101	.011	.012	.922	.376	.016	.017	.937	.523	.017	.017	.943
MICE_{INM}	60%	.101	.011	.016	.825	.369	.017	.022	.853	.530	.018	.022	.862
	70%	.100	.011	.014	.870	.371	.017	.019	.903	.529	.017	.019	.907
	80%	.100	.011	.012	.918	.373	.016	.017	.934	.527	.017	.017	.941
$\text{MICE}_{\text{INNM}}$	60%	.114	.012	.016	.750	.464	.018	.021	.009	.422	.018	.019	.000
	70%	.111	.012	.014	.827	.441	.018	.019	.058	.448	.017	.018	.010
	80%	.108	.011	.012	.893	.421	.017	.018	.265	.471	.017	.017	.132

Note. In Scenario 2 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Tables for Scenario 3 for all the MI Methods

Table F11.
The Results of the Different Robust MI Methods in Scenario 3 for $n = 50$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.096	-	.049	-	.375	-	.088	-	.529	-	.091	-
	70%	.096	-	.047	-	.375	-	.082	-	.529	-	.085	-
	80%	.096	-	.045	-	.376	-	.076	-	.527	-	.079	-
$\text{MRIC}_{R^2_{MF}}$	60%	.099	.060	.050	.927	.374	.101	.093	.939	.527	.104	.095	.944
	70%	.098	.053	.048	.912	.374	.089	.086	.936	.529	.092	.088	.933
	80%	.097	.048	.046	.913	.377	.080	.078	.943	.526	.083	.081	.942
$\text{MRIC}_{R^2_{CS}}$	60%	.100	.060	.051	.926	.374	.100	.091	.942	.526	.104	.095	.944
	70%	.098	.053	.048	.919	.375	.089	.085	.941	.527	.092	.088	.941
	80%	.097	.048	.046	.916	.376	.080	.078	.945	.527	.083	.081	.944
$\text{MRIC}_{R^2_{MZ}}$	60%	.099	.060	.051	.930	.374	.100	.093	.940	.527	.104	.096	.940
	70%	.098	.053	.048	.913	.375	.089	.085	.937	.528	.092	.088	.944
	80%	.097	.048	.046	.907	.376	.080	.078	.943	.527	.083	.081	.943
$\text{MRIC}_{R^2_N}$	60%	.100	.061	.051	.928	.374	.101	.092	.937	.526	.104	.095	.940
	70%	.097	.053	.048	.916	.375	.089	.085	.933	.528	.091	.089	.938
	80%	.097	.048	.046	.903	.376	.080	.078	.943	.527	.083	.081	.943
MRIC_{AIC}	60%	.099	.059	.051	.922	.375	.098	.092	.937	.526	.102	.095	.937
	70%	.097	.052	.048	.912	.375	.087	.084	.936	.528	.090	.088	.942
	80%	.098	.048	.046	.907	.376	.080	.078	.943	.526	.082	.081	.940
MRIC_{HL}	60%	.100	.068	.053	.919	.374	.112	.095	.947	.526	.116	.098	.948
	70%	.098	.058	.050	.908	.375	.097	.087	.944	.528	.101	.090	.949
	80%	.097	.050	.047	.902	.376	.085	.079	.948	.527	.087	.082	.951
MRNNMI	60%	.100	.061	.051	.927	.374	.100	.093	.940	.526	.103	.096	.944
	70%	.098	.053	.048	.918	.375	.089	.084	.943	.528	.092	.088	.940
	80%	.097	.048	.046	.915	.375	.080	.078	.942	.527	.083	.081	.941
DRNNMI	60%	.100	.060	.051	.922	.376	.099	.092	.936	.524	.103	.095	.938
	70%	.098	.053	.048	.919	.376	.089	.085	.937	.526	.092	.088	.941
	80%	.098	.048	.046	.912	.377	.080	.078	.942	.525	.083	.081	.945
MRPMMI	60%	.100	.061	.051	.927	.374	.102	.094	.937	.527	.106	.097	.941
	70%	.098	.054	.048	.914	.374	.090	.085	.940	.528	.093	.089	.940
	80%	.098	.049	.046	.912	.376	.081	.078	.945	.526	.083	.081	.947
$\text{MICE}_{\text{default}}$	60%	.163	.060	.076	.813	.362	.074	.081	.905	.475	.076	.079	.885
	70%	.136	.054	.063	.896	.369	.073	.077	.925	.496	.075	.076	.923
	80%	.120	.050	.053	.927	.373	.072	.073	.936	.507	.074	.074	.936
MICE_{CM}	60%	.154	.058	.076	.837	.364	.075	.084	.902	.483	.077	.082	.886
	70%	.131	.053	.062	.898	.369	.073	.078	.917	.500	.075	.079	.917
	80%	.117	.049	.053	.925	.373	.072	.074	.934	.509	.074	.075	.937
MICE_{INM}	60%	.151	.058	.067	.875	.358	.074	.078	.912	.491	.077	.081	.907
	70%	.129	.053	.058	.925	.365	.073	.076	.927	.506	.076	.078	.929
	80%	.116	.049	.051	.930	.370	.072	.073	.931	.514	.074	.075	.935
$\text{MICE}_{\text{INNM}}$	60%	.138	.057	.067	.887	.446	.079	.089	.818	.416	.077	.079	.686
	70%	.124	.053	.059	.915	.432	.077	.083	.873	.444	.075	.077	.791
	80%	.115	.049	.053	.920	.417	.075	.077	.909	.468	.074	.075	.872

Note. In Scenario 3 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Table F12.

The Results of the Different Robust MI Methods in Scenario 3 for $n = 100$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.090	-	.037	-	.376	-	.063	-	.534	-	.066	-
	70%	.093	-	.034	-	.376	-	.057	-	.532	-	.060	-
	80%	.095	-	.033	-	.377	-	.054	-	.528	-	.056	-
$\text{MRIC}_{R^2_{MF}}$	60%	.091	.043	.039	.883	.375	.075	.067	.947	.534	.077	.070	.945
	70%	.093	.038	.036	.889	.376	.064	.060	.952	.531	.066	.063	.952
	80%	.095	.034	.034	.902	.377	.058	.055	.948	.528	.059	.058	.945
$\text{MRIC}_{R^2_{CS}}$	60%	.091	.043	.039	.881	.376	.074	.067	.950	.533	.077	.070	.946
	70%	.093	.038	.036	.900	.376	.065	.060	.952	.531	.067	.062	.952
	80%	.095	.034	.034	.899	.377	.058	.056	.948	.528	.059	.058	.945
$\text{MRIC}_{R^2_{MZ}}$	60%	.091	.043	.039	.880	.375	.075	.068	.948	.534	.077	.070	.943
	70%	.093	.038	.036	.894	.376	.065	.060	.949	.531	.067	.063	.945
	80%	.095	.034	.034	.903	.377	.058	.056	.947	.528	.059	.058	.950
$\text{MRIC}_{R^2_{N}}$	60%	.091	.043	.039	.878	.376	.075	.067	.950	.534	.077	.070	.949
	70%	.093	.038	.036	.891	.376	.065	.060	.948	.531	.067	.062	.949
	80%	.095	.034	.033	.903	.377	.057	.056	.946	.528	.059	.058	.945
MRIC_{AIC}	60%	.091	.042	.039	.886	.376	.074	.067	.948	.534	.076	.070	.944
	70%	.093	.038	.036	.899	.376	.064	.060	.947	.531	.066	.062	.952
	80%	.095	.034	.034	.905	.377	.057	.055	.948	.528	.059	.058	.948
MRIC_{HL}	60%	.091	.049	.040	.883	.376	.086	.070	.960	.533	.089	.073	.959
	70%	.093	.041	.037	.897	.376	.071	.061	.963	.531	.073	.063	.965
	80%	.095	.035	.034	.905	.377	.060	.056	.955	.528	.062	.058	.956
MRNNMI	60%	.091	.043	.038	.889	.376	.075	.067	.948	.533	.077	.070	.945
	70%	.093	.038	.036	.891	.376	.065	.060	.949	.531	.067	.063	.947
	80%	.095	.034	.033	.903	.377	.058	.055	.947	.528	.059	.057	.949
DRNNMI	60%	.091	.044	.039	.876	.377	.077	.068	.952	.532	.080	.070	.953
	70%	.093	.039	.036	.897	.377	.067	.061	.949	.531	.070	.063	.952
	80%	.095	.035	.034	.901	.378	.060	.056	.957	.527	.061	.058	.955
MRPMMI	60%	.091	.045	.040	.884	.375	.078	.068	.952	.533	.080	.071	.948
	70%	.093	.039	.036	.895	.376	.067	.060	.955	.531	.069	.063	.949
	80%	.095	.035	.034	.904	.377	.059	.056	.954	.528	.061	.058	.951
$\text{MICE}_{\text{default}}$	60%	.133	.040	.057	.806	.369	.054	.065	.884	.498	.055	.064	.881
	70%	.118	.037	.045	.887	.374	.053	.059	.911	.508	.054	.058	.915
	80%	.110	.034	.038	.913	.376	.051	.054	.927	.514	.053	.055	.924
MICE_{CM}	60%	.128	.039	.056	.814	.370	.054	.067	.876	.502	.055	.066	.874
	70%	.115	.036	.044	.881	.374	.053	.059	.910	.511	.054	.060	.912
	80%	.109	.034	.038	.912	.376	.051	.055	.932	.516	.053	.055	.928
MICE_{INM}	60%	.126	.039	.052	.849	.364	.054	.063	.884	.510	.055	.064	.890
	70%	.114	.036	.043	.895	.369	.053	.057	.912	.517	.054	.059	.925
	80%	.108	.034	.037	.915	.373	.051	.054	.934	.520	.053	.055	.930
$\text{MICE}_{\text{INNM}}$	60%	.123	.039	.050	.863	.458	.057	.066	.676	.420	.055	.059	.522
	70%	.115	.037	.043	.899	.439	.055	.059	.780	.446	.054	.056	.680
	80%	.111	.035	.038	.913	.420	.053	.056	.859	.469	.053	.054	.810

Note. In Scenario 3 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Table F13.

The Results of the Different Robust MI Methods in Scenario 3 for $n = 200$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.090	-	.026	-	.375	-	.044	-	.535	-	.046	-
	70%	.092	-	.024	-	.376	-	.041	-	.532	-	.042	-
	80%	.095	-	.023	-	.376	-	.038	-	.529	-	.040	-
$\text{MRIC}_{R^2_{MF}}$	60%	.090	.031	.028	.890	.375	.054	.048	.950	.535	.056	.049	.949
	70%	.092	.027	.025	.907	.376	.046	.043	.958	.532	.048	.044	.954
	80%	.094	.025	.024	.912	.376	.041	.040	.949	.530	.042	.041	.949
$\text{MRIC}_{R^2_{CS}}$	60%	.090	.031	.028	.888	.375	.054	.048	.951	.534	.055	.049	.949
	70%	.092	.027	.026	.904	.376	.046	.043	.954	.532	.048	.044	.954
	80%	.095	.025	.024	.916	.376	.041	.040	.951	.529	.042	.041	.947
$\text{MRIC}_{R^2_{MZ}}$	60%	.090	.032	.028	.890	.375	.054	.048	.950	.535	.056	.050	.948
	70%	.092	.027	.026	.913	.376	.046	.042	.954	.532	.048	.044	.951
	80%	.095	.025	.024	.911	.376	.041	.039	.951	.529	.042	.041	.947
$\text{MRIC}_{R^2_{N}}$	60%	.090	.032	.028	.893	.376	.055	.048	.951	.534	.056	.049	.949
	70%	.092	.027	.025	.903	.376	.047	.043	.955	.532	.048	.044	.954
	80%	.095	.025	.024	.912	.376	.041	.040	.950	.529	.042	.041	.948
MRIC_{AIC}	60%	.090	.031	.028	.889	.375	.053	.047	.951	.534	.055	.049	.950
	70%	.092	.027	.025	.910	.376	.046	.043	.955	.532	.048	.044	.951
	80%	.095	.025	.024	.915	.376	.041	.039	.949	.529	.042	.041	.946
MRIC_{HL}	60%	.090	.033	.028	.896	.376	.058	.049	.957	.534	.060	.050	.958
	70%	.092	.028	.026	.911	.376	.048	.043	.961	.532	.050	.044	.957
	80%	.095	.025	.024	.914	.376	.042	.040	.952	.530	.043	.041	.946
MRNNMI	60%	.090	.031	.028	.892	.376	.054	.048	.953	.534	.056	.049	.953
	70%	.092	.028	.026	.901	.376	.046	.043	.955	.532	.048	.044	.950
	80%	.095	.025	.024	.912	.376	.041	.040	.950	.530	.042	.041	.949
DRNNMI	60%	.091	.035	.029	.892	.376	.060	.049	.962	.534	.062	.050	.964
	70%	.092	.030	.026	.906	.376	.051	.044	.963	.532	.053	.045	.964
	80%	.095	.026	.024	.912	.376	.044	.040	.956	.529	.045	.041	.956
MRPMMI	60%	.090	.034	.029	.900	.376	.060	.049	.958	.534	.062	.051	.957
	70%	.092	.030	.026	.912	.376	.051	.044	.962	.532	.052	.045	.960
	80%	.095	.026	.024	.920	.376	.044	.040	.954	.530	.045	.041	.956
$\text{MICE}_{\text{default}}$	60%	.117	.027	.039	.819	.373	.039	.048	.876	.511	.039	.047	.883
	70%	.109	.025	.031	.884	.375	.038	.042	.911	.516	.038	.042	.919
	80%	.105	.024	.027	.913	.375	.037	.038	.937	.520	.037	.039	.933
MICE_{CM}	60%	.114	.027	.039	.820	.373	.039	.048	.876	.513	.039	.047	.886
	70%	.108	.025	.031	.877	.375	.038	.042	.913	.517	.038	.042	.918
	80%	.104	.024	.027	.910	.375	.037	.039	.936	.520	.037	.039	.934
MICE_{INM}	60%	.113	.027	.037	.832	.366	.038	.047	.868	.521	.039	.047	.895
	70%	.106	.025	.030	.887	.370	.037	.042	.913	.523	.038	.042	.924
	80%	.103	.024	.026	.914	.372	.036	.038	.937	.524	.037	.039	.935
$\text{MICE}_{\text{INNM}}$	60%	.117	.028	.035	.852	.461	.041	.047	.445	.421	.039	.042	.270
	70%	.113	.026	.030	.901	.440	.039	.042	.622	.447	.038	.039	.487
	80%	.109	.025	.027	.915	.420	.038	.039	.778	.471	.037	.038	.709

Note. In Scenario 3 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Table F14.

The Results of the Different Robust MI Methods in Scenario 3 for $n = 500$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.090	-	.016	-	.375	-	.027	-	.535	-	.029	-
	70%	.092	-	.015	-	.375	-	.026	-	.533	-	.027	-
	80%	.094	-	.014	-	.376	-	.024	-	.530	-	.025	-
$\text{MRIC}_{R^2_{MF}}$	60%	.090	.020	.018	.892	.375	.035	.030	.960	.535	.036	.031	.950
	70%	.092	.018	.016	.904	.375	.030	.027	.958	.533	.031	.028	.945
	80%	.094	.016	.015	.913	.376	.026	.025	.957	.530	.027	.026	.948
$\text{MRIC}_{R^2_{CS}}$	60%	.090	.020	.018	.889	.375	.035	.030	.959	.536	.036	.031	.944
	70%	.092	.018	.016	.899	.375	.030	.027	.956	.533	.031	.028	.951
	80%	.094	.016	.015	.914	.376	.026	.025	.953	.530	.027	.026	.944
$\text{MRIC}_{R^2_{MZ}}$	60%	.090	.021	.018	.889	.375	.035	.030	.960	.536	.036	.031	.944
	70%	.092	.018	.016	.905	.375	.030	.027	.953	.533	.031	.028	.946
	80%	.094	.016	.015	.919	.375	.026	.025	.956	.530	.027	.026	.945
$\text{MRIC}_{R^2_{N}}$	60%	.090	.020	.018	.889	.375	.035	.030	.956	.536	.036	.031	.944
	70%	.092	.018	.016	.906	.375	.030	.027	.958	.533	.031	.028	.948
	80%	.094	.016	.015	.912	.376	.026	.025	.956	.530	.027	.026	.946
MRIC_{AIC}	60%	.090	.020	.018	.883	.375	.035	.030	.961	.536	.036	.031	.944
	70%	.092	.018	.016	.902	.375	.030	.027	.952	.533	.031	.028	.942
	80%	.094	.016	.015	.913	.376	.026	.025	.955	.530	.027	.026	.943
MRIC_{HL}	60%	.090	.021	.018	.893	.375	.036	.030	.967	.536	.038	.031	.950
	70%	.092	.018	.016	.907	.375	.031	.027	.962	.533	.032	.028	.954
	80%	.094	.016	.015	.920	.376	.027	.025	.956	.530	.028	.026	.949
MRNNMI	60%	.090	.021	.018	.891	.375	.035	.030	.955	.535	.036	.031	.948
	70%	.092	.018	.016	.899	.375	.030	.027	.958	.533	.031	.028	.943
	80%	.094	.016	.015	.916	.376	.026	.025	.959	.530	.027	.026	.947
DRNNMI	60%	.090	.025	.019	.894	.374	.043	.032	.974	.536	.045	.033	.966
	70%	.092	.021	.017	.910	.375	.036	.028	.971	.533	.037	.029	.960
	80%	.094	.018	.015	.926	.376	.030	.025	.970	.530	.031	.027	.959
MRPMMI	60%	.090	.027	.019	.904	.375	.046	.032	.979	.535	.047	.034	.974
	70%	.092	.022	.017	.924	.375	.037	.029	.972	.533	.038	.029	.967
	80%	.094	.018	.015	.927	.376	.030	.026	.967	.530	.031	.027	.956
$\text{MICE}_{\text{default}}$	60%	.106	.017	.024	.814	.374	.025	.031	.874	.519	.025	.030	.886
	70%	.104	.016	.020	.878	.375	.024	.027	.911	.521	.024	.027	.917
	80%	.102	.015	.017	.915	.376	.023	.024	.938	.522	.024	.025	.928
MICE_{CM}	60%	.106	.017	.024	.811	.375	.025	.031	.867	.520	.025	.031	.883
	70%	.103	.016	.020	.877	.375	.024	.027	.914	.522	.024	.027	.918
	80%	.102	.015	.017	.909	.376	.023	.024	.935	.522	.024	.025	.930
MICE_{INM}	60%	.103	.016	.023	.829	.368	.024	.031	.863	.528	.025	.030	.881
	70%	.102	.016	.019	.883	.371	.024	.027	.912	.528	.024	.027	.915
	80%	.101	.015	.017	.914	.373	.023	.024	.934	.526	.024	.025	.929
$\text{MICE}_{\text{INNM}}$	60%	.115	.018	.022	.823	.463	.026	.030	.112	.422	.025	.026	.027
	70%	.111	.017	.019	.875	.441	.025	.027	.279	.448	.024	.025	.134
	80%	.108	.016	.017	.917	.421	.024	.025	.547	.472	.024	.024	.418

Note. In Scenario 3 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Table F15.

The Results of the Different Robust MI Methods in Scenario 3 for $n = 1000$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.090	-	.012	-	.374	-	.020	-	.536	-	.020	-
	70%	.092	-	.011	-	.375	-	.018	-	.533	-	.019	-
	80%	.095	-	.010	-	.375	-	.017	-	.530	-	.017	-
$MRIC_{R^2_{MF}}$	60%	.090	.015	.013	.870	.374	.025	.021	.959	.535	.026	.022	.935
	70%	.092	.013	.012	.882	.375	.021	.019	.962	.533	.022	.020	.939
	80%	.094	.011	.011	.899	.376	.019	.018	.959	.530	.019	.018	.945
$MRIC_{R^2_{CS}}$	60%	.090	.015	.013	.866	.374	.025	.021	.961	.536	.026	.022	.941
	70%	.092	.013	.012	.881	.375	.021	.019	.959	.533	.022	.020	.943
	80%	.095	.011	.011	.909	.375	.019	.018	.960	.530	.019	.018	.949
$MRIC_{R^2_{MZ}}$	60%	.090	.015	.013	.861	.374	.025	.021	.960	.536	.026	.022	.936
	70%	.092	.013	.012	.886	.375	.021	.019	.959	.533	.022	.020	.936
	80%	.094	.011	.011	.907	.375	.019	.018	.957	.530	.019	.018	.947
$MRIC_{R^2_N}$	60%	.090	.015	.013	.865	.374	.025	.021	.956	.535	.026	.022	.938
	70%	.092	.013	.012	.881	.375	.021	.019	.960	.533	.022	.020	.938
	80%	.094	.011	.011	.907	.376	.019	.017	.958	.530	.019	.018	.947
$MRIC_{AIC}$	60%	.090	.015	.013	.858	.374	.025	.021	.961	.536	.026	.022	.934
	70%	.092	.013	.012	.880	.375	.021	.019	.961	.533	.022	.020	.938
	80%	.094	.011	.011	.900	.375	.019	.018	.958	.530	.019	.018	.945
$MRIC_{HL}$	60%	.090	.016	.013	.867	.374	.027	.022	.965	.536	.028	.023	.949
	70%	.092	.013	.012	.890	.375	.022	.019	.967	.533	.023	.020	.946
	80%	.094	.012	.011	.909	.376	.019	.018	.964	.530	.020	.018	.952
MRNNMI	60%	.090	.015	.013	.862	.375	.025	.021	.964	.535	.026	.022	.941
	70%	.092	.013	.012	.880	.375	.022	.019	.958	.533	.022	.020	.942
	80%	.095	.011	.011	.905	.376	.019	.018	.960	.530	.019	.018	.946
DRNNMI	60%	.090	.019	.014	.884	.374	.033	.023	.978	.535	.035	.024	.963
	70%	.092	.016	.012	.896	.375	.027	.020	.977	.533	.028	.021	.966
	80%	.095	.013	.011	.918	.375	.022	.018	.969	.530	.023	.019	.966
MRPMMI	60%	.090	.023	.015	.903	.374	.039	.025	.979	.536	.040	.026	.972
	70%	.092	.018	.013	.918	.375	.031	.021	.985	.533	.032	.022	.976
	80%	.094	.014	.011	.930	.376	.024	.019	.977	.530	.025	.019	.974
$MICE_{default}$	60%	.104	.012	.017	.812	.375	.017	.022	.871	.521	.018	.022	.880
	70%	.102	.011	.014	.876	.375	.017	.019	.912	.523	.017	.019	.915
	80%	.101	.011	.012	.922	.376	.016	.017	.939	.523	.017	.017	.941
$MICE_{CM}$	60%	.103	.012	.017	.811	.375	.017	.022	.873	.522	.018	.022	.880
	70%	.102	.011	.014	.872	.375	.017	.019	.911	.523	.017	.019	.918
	80%	.101	.011	.012	.922	.376	.016	.017	.937	.523	.017	.017	.943
$MICE_{INM}$	60%	.101	.011	.016	.825	.369	.017	.022	.853	.530	.018	.022	.862
	70%	.100	.011	.014	.870	.371	.017	.019	.903	.529	.017	.019	.907
	80%	.100	.011	.012	.918	.373	.016	.017	.934	.527	.017	.017	.941
$MICE_{INNM}$	60%	.114	.012	.016	.750	.464	.018	.021	.009	.422	.018	.019	.000
	70%	.111	.012	.014	.827	.441	.018	.019	.058	.448	.017	.018	.010
	80%	.108	.011	.012	.893	.421	.017	.018	.265	.471	.017	.017	.132

Note. In Scenario 3 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Tables for Scenario 4 for all the MI Methods

Table F16.
The Results of the Different Robust MI Methods in Scenario 4 for $n = 50$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.097	-	.049	-	.374	-	.088	-	.529	-	.091	-
	70%	.096	-	.047	-	.375	-	.082	-	.529	-	.085	-
	80%	.097	-	.045	-	.376	-	.076	-	.527	-	.079	-
$\text{MRIC}_{R^2_{MF}}$	60%	.099	.060	.051	.924	.373	.100	.092	.937	.528	.103	.096	.940
	70%	.098	.053	.048	.919	.374	.088	.085	.936	.528	.091	.088	.940
	80%	.097	.048	.046	.913	.375	.080	.078	.940	.527	.083	.081	.935
$\text{MRIC}_{R^2_{CS}}$	60%	.100	.060	.051	.930	.373	.099	.092	.938	.527	.103	.096	.940
	70%	.098	.053	.048	.916	.374	.088	.085	.939	.529	.092	.089	.935
	80%	.098	.048	.046	.915	.376	.080	.078	.941	.527	.083	.081	.942
$\text{MRIC}_{R^2_{MZ}}$	60%	.100	.061	.050	.926	.373	.100	.092	.935	.527	.104	.095	.943
	70%	.098	.053	.048	.915	.374	.088	.085	.938	.529	.091	.088	.943
	80%	.098	.048	.046	.913	.375	.080	.077	.945	.527	.083	.081	.945
$\text{MRIC}_{R^2_{N}}$	60%	.100	.061	.051	.928	.374	.100	.092	.941	.526	.104	.095	.941
	70%	.098	.053	.048	.916	.374	.088	.085	.938	.528	.091	.088	.941
	80%	.098	.048	.046	.908	.376	.080	.078	.945	.527	.082	.081	.945
MRIC_{AIC}	60%	.099	.059	.050	.922	.374	.098	.092	.939	.527	.101	.096	.937
	70%	.098	.052	.048	.915	.374	.087	.084	.937	.528	.090	.088	.938
	80%	.098	.048	.046	.912	.375	.079	.078	.941	.526	.082	.081	.943
MRIC_{HL}	60%	.100	.069	.053	.915	.374	.114	.095	.945	.526	.118	.098	.950
	70%	.098	.058	.049	.910	.374	.098	.087	.948	.528	.101	.090	.948
	80%	.098	.051	.047	.904	.375	.085	.078	.948	.527	.088	.081	.949
MRNNMI	60%	.100	.060	.051	.928	.374	.100	.093	.938	.526	.103	.096	.946
	70%	.098	.053	.049	.918	.374	.088	.084	.939	.529	.091	.088	.941
	80%	.098	.048	.046	.910	.375	.080	.078	.947	.527	.083	.081	.942
DRNNMI	60%	.100	.057	.050	.929	.375	.095	.091	.930	.525	.098	.094	.934
	70%	.098	.051	.048	.918	.375	.086	.084	.936	.527	.088	.087	.937
	80%	.098	.047	.046	.911	.376	.079	.077	.943	.526	.081	.080	.941
MRPMMI	60%	.099	.062	.051	.922	.374	.103	.094	.938	.527	.107	.097	.940
	70%	.098	.054	.048	.912	.374	.090	.085	.945	.528	.093	.088	.942
	80%	.098	.049	.047	.908	.375	.081	.078	.946	.527	.083	.081	.945
$\text{MICE}_{\text{default}}$	60%	.164	.060	.076	.812	.362	.074	.081	.905	.475	.076	.079	.886
	70%	.136	.054	.063	.896	.369	.073	.077	.925	.495	.075	.076	.923
	80%	.121	.050	.053	.925	.372	.072	.073	.937	.507	.074	.075	.937
MICE_{CM}	60%	.154	.058	.076	.835	.363	.075	.083	.903	.482	.077	.082	.888
	70%	.131	.053	.062	.898	.369	.073	.078	.917	.500	.075	.078	.919
	80%	.118	.049	.053	.925	.373	.072	.073	.934	.509	.074	.075	.937
MICE_{INM}	60%	.151	.058	.068	.873	.358	.074	.078	.913	.491	.077	.081	.908
	70%	.129	.053	.058	.925	.365	.073	.075	.928	.506	.076	.078	.930
	80%	.117	.049	.051	.930	.369	.072	.072	.931	.514	.074	.075	.935
$\text{MICE}_{\text{INNM}}$	60%	.138	.057	.067	.887	.445	.079	.089	.820	.416	.077	.079	.686
	70%	.124	.053	.059	.916	.431	.077	.083	.873	.444	.075	.077	.792
	80%	.116	.049	.053	.920	.416	.075	.077	.910	.468	.074	.075	.872

Note. In Scenario 4 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Table F17.

The Results of the Different Robust MI Methods in Scenario 4 for $n = 100$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.091	-	.036	-	.375	-	.063	-	.534	-	.066	-
	70%	.094	-	.034	-	.375	-	.057	-	.531	-	.060	-
	80%	.096	-	.032	-	.376	-	.054	-	.528	-	.056	-
$\text{MRIC}_{R^2_{MF}}$	60%	.092	.043	.039	.885	.375	.074	.067	.946	.533	.076	.070	.946
	70%	.093	.038	.036	.897	.376	.064	.060	.953	.531	.066	.062	.950
	80%	.096	.034	.033	.910	.376	.057	.055	.948	.528	.059	.058	.947
$\text{MRIC}_{R^2_{CS}}$	60%	.092	.043	.039	.888	.375	.074	.067	.948	.533	.076	.070	.947
	70%	.094	.038	.036	.903	.375	.064	.060	.949	.531	.066	.062	.946
	80%	.096	.034	.033	.912	.377	.057	.055	.947	.528	.059	.058	.948
$\text{MRIC}_{R^2_{MZ}}$	60%	.092	.043	.039	.886	.375	.075	.068	.949	.533	.077	.070	.950
	70%	.094	.038	.036	.899	.375	.064	.060	.952	.531	.066	.062	.948
	80%	.096	.034	.033	.912	.376	.057	.055	.950	.528	.059	.057	.950
$\text{MRIC}_{R^2_N}$	60%	.091	.043	.039	.883	.375	.074	.067	.949	.533	.076	.071	.945
	70%	.094	.038	.036	.903	.375	.064	.060	.946	.531	.066	.062	.949
	80%	.096	.034	.033	.906	.376	.057	.056	.947	.528	.059	.058	.947
MRIC_{AIC}	60%	.092	.043	.039	.889	.375	.073	.067	.945	.533	.076	.070	.943
	70%	.094	.038	.036	.901	.375	.064	.059	.953	.531	.066	.062	.949
	80%	.096	.034	.033	.910	.376	.058	.056	.950	.528	.059	.058	.945
MRIC_{HL}	60%	.092	.049	.040	.889	.375	.085	.070	.961	.533	.088	.073	.960
	70%	.094	.041	.036	.902	.375	.070	.061	.960	.531	.072	.064	.959
	80%	.096	.036	.033	.909	.376	.060	.056	.956	.528	.062	.058	.952
MRNNMI	60%	.092	.043	.038	.888	.375	.074	.067	.947	.533	.077	.070	.946
	70%	.094	.038	.036	.901	.375	.064	.060	.948	.531	.066	.062	.950
	80%	.096	.034	.033	.907	.376	.057	.056	.950	.528	.059	.058	.945
DRNNMI	60%	.092	.040	.038	.884	.376	.069	.066	.941	.533	.072	.069	.943
	70%	.094	.036	.035	.903	.376	.062	.059	.944	.530	.064	.062	.946
	80%	.096	.034	.033	.910	.377	.057	.055	.948	.527	.058	.057	.950
MRPMMI	60%	.092	.045	.039	.891	.375	.078	.068	.956	.533	.081	.071	.953
	70%	.094	.039	.036	.902	.375	.067	.060	.955	.531	.069	.063	.954
	80%	.096	.035	.033	.913	.377	.059	.056	.953	.527	.061	.058	.951
$\text{MICE}_{\text{default}}$	60%	.134	.040	.057	.803	.369	.054	.065	.884	.497	.055	.064	.881
	70%	.119	.037	.045	.888	.373	.053	.059	.911	.508	.054	.058	.915
	80%	.111	.035	.038	.917	.375	.051	.054	.927	.514	.053	.055	.926
MICE_{CM}	60%	.128	.039	.057	.812	.370	.054	.067	.876	.502	.055	.065	.874
	70%	.116	.036	.044	.882	.374	.053	.059	.910	.510	.054	.060	.912
	80%	.109	.034	.038	.915	.375	.051	.055	.932	.515	.053	.055	.930
MICE_{INM}	60%	.127	.039	.052	.849	.363	.054	.063	.884	.510	.055	.064	.891
	70%	.115	.036	.042	.897	.369	.053	.057	.912	.517	.054	.059	.924
	80%	.109	.034	.037	.918	.372	.051	.054	.934	.519	.053	.055	.931
$\text{MICE}_{\text{INNM}}$	60%	.124	.039	.050	.864	.457	.057	.066	.678	.419	.055	.059	.522
	70%	.116	.037	.042	.900	.438	.055	.060	.783	.445	.054	.056	.678
	80%	.112	.035	.038	.917	.420	.053	.056	.862	.469	.053	.054	.809

Note. In Scenario 4 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Table F18.

The Results of the Different Robust MI Methods in Scenario 4 for $n = 200$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.091	-	.026	-	.375	-	.044	-	.534	-	.046	-
	70%	.093	-	.024	-	.376	-	.041	-	.532	-	.042	-
	80%	.095	-	.023	-	.376	-	.038	-	.529	-	.040	-
$\text{MRIC}_{R^2_{MF}}$	60%	.091	.031	.028	.894	.375	.053	.048	.945	.534	.055	.049	.947
	70%	.093	.027	.025	.914	.375	.046	.042	.957	.532	.047	.043	.954
	80%	.095	.024	.024	.912	.376	.041	.039	.950	.530	.042	.041	.949
$\text{MRIC}_{R^2_{CS}}$	60%	.091	.031	.028	.903	.375	.053	.047	.955	.534	.055	.049	.950
	70%	.092	.027	.025	.908	.376	.046	.043	.952	.532	.047	.044	.950
	80%	.095	.024	.024	.916	.376	.041	.040	.948	.529	.042	.041	.946
$\text{MRIC}_{R^2_{MZ}}$	60%	.091	.031	.028	.902	.376	.054	.048	.949	.534	.056	.049	.947
	70%	.093	.027	.025	.905	.375	.046	.042	.953	.532	.047	.044	.951
	80%	.095	.024	.024	.913	.376	.041	.039	.949	.529	.042	.041	.948
$\text{MRIC}_{R^2_{N}}$	60%	.090	.031	.028	.896	.375	.053	.048	.953	.535	.055	.049	.946
	70%	.093	.027	.025	.906	.376	.046	.043	.951	.532	.047	.044	.957
	80%	.095	.024	.024	.915	.376	.041	.040	.948	.529	.042	.041	.947
MRIC_{AIC}	60%	.091	.032	.028	.896	.375	.055	.048	.955	.534	.057	.050	.952
	70%	.093	.028	.025	.907	.376	.047	.043	.957	.532	.049	.044	.951
	80%	.095	.025	.024	.916	.376	.042	.040	.954	.530	.043	.041	.951
MRIC_{HL}	60%	.091	.033	.028	.909	.376	.058	.049	.954	.534	.059	.050	.955
	70%	.093	.028	.026	.910	.375	.048	.043	.955	.532	.049	.044	.955
	80%	.095	.025	.024	.920	.376	.042	.040	.952	.529	.043	.041	.948
MRNNMI	60%	.091	.031	.028	.893	.375	.054	.047	.953	.534	.055	.048	.952
	70%	.093	.027	.025	.910	.375	.046	.043	.956	.532	.047	.044	.952
	80%	.095	.024	.024	.914	.376	.041	.039	.947	.530	.042	.041	.949
DRNNMI	60%	.091	.029	.027	.894	.375	.050	.047	.943	.534	.051	.048	.941
	70%	.093	.026	.025	.906	.376	.044	.042	.954	.532	.046	.043	.952
	80%	.095	.024	.024	.913	.376	.040	.039	.948	.529	.041	.041	.946
MRPMMI	60%	.090	.035	.028	.908	.375	.060	.049	.963	.535	.062	.050	.959
	70%	.093	.030	.026	.916	.376	.051	.044	.961	.532	.053	.045	.962
	80%	.095	.026	.024	.919	.376	.043	.040	.957	.529	.045	.042	.952
$\text{MICE}_{\text{default}}$	60%	.117	.027	.039	.819	.372	.039	.048	.876	.511	.039	.047	.883
	70%	.109	.025	.031	.888	.375	.038	.042	.911	.516	.038	.042	.919
	80%	.105	.024	.027	.914	.375	.037	.038	.936	.520	.037	.039	.933
MICE_{CM}	60%	.115	.027	.039	.821	.373	.039	.048	.875	.513	.039	.047	.886
	70%	.108	.025	.031	.881	.375	.038	.042	.914	.517	.038	.042	.918
	80%	.105	.024	.027	.910	.375	.037	.039	.935	.520	.037	.039	.934
MICE_{INM}	60%	.113	.027	.037	.834	.366	.038	.047	.868	.521	.039	.047	.897
	70%	.107	.025	.030	.892	.370	.037	.042	.913	.523	.038	.042	.925
	80%	.103	.024	.026	.915	.372	.036	.038	.937	.524	.037	.039	.935
$\text{MICE}_{\text{INNM}}$	60%	.118	.028	.035	.853	.461	.041	.047	.449	.421	.039	.042	.269
	70%	.113	.026	.030	.903	.440	.039	.042	.624	.447	.038	.039	.486
	80%	.109	.025	.027	.916	.420	.038	.039	.779	.471	.037	.038	.708

Note. In Scenario 4 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Table F19.

The Results of the Different Robust MI Methods in Scenario 4 for $n = 500$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.090	-	.016	-	.375	-	.027	-	.535	-	.029	-
	70%	.092	-	.015	-	.375	-	.026	-	.533	-	.027	-
	80%	.094	-	.014	-	.376	-	.024	-	.530	-	.025	-
$\text{MRIC}_{R^2_{MF}}$	60%	.090	.020	.017	.887	.374	.034	.029	.956	.536	.035	.031	.945
	70%	.092	.017	.016	.899	.375	.029	.027	.954	.533	.030	.028	.942
	80%	.094	.015	.015	.904	.376	.026	.025	.953	.530	.027	.026	.946
$\text{MRIC}_{R^2_{CS}}$	60%	.090	.020	.018	.881	.375	.034	.030	.955	.535	.035	.031	.941
	70%	.092	.017	.016	.901	.375	.029	.027	.953	.533	.030	.028	.943
	80%	.094	.016	.015	.915	.376	.026	.025	.953	.530	.027	.026	.943
$\text{MRIC}_{R^2_{MZ}}$	60%	.090	.020	.017	.887	.375	.034	.030	.956	.535	.035	.031	.940
	70%	.092	.017	.016	.899	.375	.029	.027	.952	.533	.030	.028	.944
	80%	.094	.016	.015	.914	.376	.026	.025	.953	.530	.027	.026	.943
$\text{MRIC}_{R^2_{N}}$	60%	.090	.020	.017	.880	.375	.034	.029	.956	.535	.035	.031	.945
	70%	.092	.017	.016	.900	.375	.029	.027	.953	.533	.030	.028	.945
	80%	.094	.016	.015	.911	.376	.026	.025	.952	.530	.027	.026	.943
MRIC_{AIC}	60%	.090	.022	.018	.895	.374	.038	.031	.963	.536	.040	.032	.961
	70%	.092	.019	.016	.903	.375	.032	.027	.967	.533	.033	.028	.955
	80%	.094	.016	.015	.916	.376	.028	.025	.960	.530	.028	.026	.953
MRIC_{HL}	60%	.090	.020	.017	.893	.375	.035	.030	.963	.536	.036	.031	.950
	70%	.092	.018	.016	.902	.375	.030	.027	.958	.533	.030	.028	.944
	80%	.094	.016	.015	.916	.376	.026	.025	.958	.530	.027	.026	.945
MRNNMI	60%	.090	.020	.017	.887	.375	.034	.030	.959	.535	.035	.031	.945
	70%	.092	.017	.016	.900	.375	.029	.027	.953	.533	.030	.028	.947
	80%	.094	.016	.015	.909	.376	.026	.025	.951	.530	.027	.026	.945
DRNNMI	60%	.090	.019	.018	.875	.375	.032	.029	.953	.535	.033	.031	.936
	70%	.092	.017	.016	.894	.375	.028	.027	.950	.533	.029	.028	.941
	80%	.094	.015	.015	.908	.376	.025	.025	.951	.530	.026	.026	.939
MRPMMI	60%	.090	.026	.019	.907	.375	.046	.033	.977	.535	.047	.034	.970
	70%	.092	.022	.017	.915	.375	.037	.029	.974	.533	.038	.030	.967
	80%	.094	.018	.015	.928	.376	.030	.025	.970	.530	.031	.026	.960
$\text{MICE}_{\text{default}}$	60%	.106	.017	.024	.814	.374	.025	.031	.874	.519	.025	.030	.886
	70%	.104	.016	.020	.878	.375	.024	.027	.911	.521	.024	.027	.917
	80%	.102	.015	.017	.915	.376	.023	.024	.938	.522	.024	.025	.928
MICE_{CM}	60%	.106	.017	.024	.811	.375	.025	.031	.867	.520	.025	.031	.883
	70%	.103	.016	.020	.877	.375	.024	.027	.914	.522	.024	.027	.918
	80%	.102	.015	.017	.909	.376	.023	.024	.935	.522	.024	.025	.930
MICE_{INM}	60%	.103	.016	.023	.829	.368	.024	.031	.863	.528	.025	.030	.881
	70%	.102	.016	.019	.883	.371	.024	.027	.912	.528	.024	.027	.915
	80%	.101	.015	.017	.914	.373	.023	.024	.934	.526	.024	.025	.929
$\text{MICE}_{\text{INNM}}$	60%	.115	.018	.022	.823	.463	.026	.030	.112	.422	.025	.026	.027
	70%	.111	.017	.019	.875	.441	.025	.027	.279	.448	.024	.025	.134
	80%	.108	.016	.017	.917	.421	.024	.025	.547	.472	.024	.024	.418

Note. In Scenario 4 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Table F20.

The Results of the Different Robust MI Methods in Scenario 4 for $n = 1000$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.090	-	.012	-	.374	-	.020	-	.536	-	.020	-
	70%	.092	-	.011	-	.375	-	.018	-	.533	-	.019	-
	80%	.095	-	.010	-	.375	-	.017	-	.530	-	.017	-
$MRIC_{R^2_{MF}}$	60%	.090	.014	.012	.863	.374	.024	.021	.959	.535	.024	.022	.929
	70%	.092	.012	.012	.872	.375	.021	.019	.959	.533	.021	.020	.937
	80%	.094	.011	.011	.903	.376	.018	.017	.957	.530	.019	.018	.946
$MRIC_{R^2_{CS}}$	60%	.090	.014	.013	.857	.374	.024	.021	.955	.536	.025	.022	.932
	70%	.092	.012	.011	.875	.375	.020	.019	.956	.533	.021	.020	.937
	80%	.095	.011	.011	.905	.375	.018	.017	.956	.530	.019	.018	.941
$MRIC_{R^2_{MZ}}$	60%	.090	.014	.013	.860	.375	.024	.021	.957	.535	.025	.022	.929
	70%	.092	.012	.011	.881	.375	.021	.019	.957	.533	.021	.020	.931
	80%	.095	.011	.011	.900	.375	.018	.017	.959	.530	.019	.018	.943
$MRIC_{R^2_{N}}$	60%	.090	.014	.013	.857	.374	.024	.021	.956	.535	.025	.022	.934
	70%	.092	.012	.012	.873	.375	.020	.019	.957	.533	.021	.020	.933
	80%	.095	.011	.011	.897	.375	.018	.017	.955	.530	.019	.018	.942
$MRIC_{AIC}$	60%	.090	.017	.013	.870	.375	.029	.023	.968	.535	.030	.023	.951
	70%	.092	.014	.012	.890	.375	.024	.020	.967	.533	.025	.020	.953
	80%	.094	.012	.011	.908	.376	.020	.018	.967	.530	.021	.018	.956
$MRIC_{HL}$	60%	.090	.014	.013	.865	.375	.025	.021	.961	.535	.025	.022	.935
	70%	.092	.012	.012	.879	.375	.021	.019	.961	.533	.022	.020	.938
	80%	.094	.011	.010	.905	.375	.019	.018	.960	.530	.019	.018	.945
MRNNMI	60%	.090	.014	.013	.855	.375	.024	.021	.956	.535	.025	.022	.932
	70%	.092	.012	.011	.877	.375	.021	.019	.958	.533	.021	.020	.940
	80%	.095	.011	.011	.902	.375	.018	.017	.956	.530	.019	.018	.941
DRNNMI	60%	.090	.013	.012	.845	.374	.023	.021	.951	.536	.023	.022	.922
	70%	.092	.012	.012	.867	.375	.020	.019	.954	.533	.021	.020	.928
	80%	.095	.011	.010	.899	.375	.018	.017	.956	.530	.019	.018	.940
MRPMMI	60%	.090	.022	.015	.904	.375	.039	.025	.980	.536	.040	.026	.973
	70%	.092	.018	.013	.918	.375	.031	.021	.980	.533	.032	.022	.971
	80%	.094	.014	.011	.928	.376	.024	.019	.978	.530	.024	.019	.971
$MICE_{default}$	60%	.104	.012	.017	.812	.375	.017	.022	.871	.521	.018	.022	.880
	70%	.102	.011	.014	.876	.375	.017	.019	.912	.523	.017	.019	.915
	80%	.101	.011	.012	.922	.376	.016	.017	.939	.523	.017	.017	.941
$MICE_{CM}$	60%	.103	.012	.017	.811	.375	.017	.022	.873	.522	.018	.022	.880
	70%	.102	.011	.014	.872	.375	.017	.019	.911	.523	.017	.019	.918
	80%	.101	.011	.012	.922	.376	.016	.017	.937	.523	.017	.017	.943
$MICE_{INM}$	60%	.101	.011	.016	.825	.369	.017	.022	.853	.530	.018	.022	.862
	70%	.100	.011	.014	.870	.371	.017	.019	.903	.529	.017	.019	.907
	80%	.100	.011	.012	.918	.373	.016	.017	.934	.527	.017	.017	.941
$MICE_{INNM}$	60%	.114	.012	.016	.750	.464	.018	.021	.009	.422	.018	.019	.000
	70%	.111	.012	.014	.827	.441	.018	.019	.058	.448	.017	.018	.010
	80%	.108	.011	.012	.893	.421	.017	.018	.265	.471	.017	.017	.132

Note. In Scenario 4 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Tables for Scenario 5 for all the MI Methods

Table F21.

The Results of the Different Robust MI Methods in Scenario 5 for $n = 50$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.096	-	.049	-	.375	-	.088	-	.529	-	.091	-
	70%	.096	-	.047	-	.375	-	.082	-	.529	-	.085	-
	80%	.096	-	.045	-	.376	-	.076	-	.527	-	.079	-
$\text{MRIC}_{R^2_{MF}}$	60%	.099	.060	.050	.929	.374	.100	.093	.939	.527	.103	.096	.943
	70%	.097	.053	.048	.908	.374	.089	.086	.936	.528	.092	.088	.943
	80%	.097	.048	.046	.907	.376	.080	.078	.942	.527	.083	.081	.942
$\text{MRIC}_{R^2_{CS}}$	60%	.100	.060	.051	.931	.374	.100	.092	.941	.526	.103	.095	.943
	70%	.098	.053	.048	.918	.375	.089	.085	.940	.528	.092	.089	.937
	80%	.097	.048	.046	.910	.376	.080	.078	.944	.527	.083	.081	.942
$\text{MRIC}_{R^2_{MZ}}$	60%	.099	.060	.050	.926	.374	.100	.093	.936	.527	.103	.095	.941
	70%	.098	.053	.048	.910	.374	.089	.085	.938	.528	.092	.088	.940
	80%	.097	.048	.046	.914	.376	.080	.078	.944	.527	.083	.081	.942
$\text{MRIC}_{R^2_{N}}$	60%	.100	.061	.052	.921	.374	.100	.092	.935	.526	.104	.095	.942
	70%	.097	.053	.048	.918	.374	.089	.086	.939	.528	.092	.089	.936
	80%	.097	.048	.046	.909	.376	.080	.078	.941	.527	.083	.081	.946
MRIC_{AIC}	60%	.099	.059	.050	.926	.375	.098	.092	.937	.526	.101	.095	.941
	70%	.097	.052	.048	.911	.375	.087	.085	.940	.528	.090	.088	.935
	80%	.097	.047	.046	.907	.376	.079	.078	.943	.526	.082	.081	.940
MRIC_{HL}	60%	.100	.068	.053	.921	.374	.112	.095	.948	.526	.116	.098	.951
	70%	.098	.058	.050	.904	.375	.097	.087	.944	.528	.100	.090	.946
	80%	.097	.050	.046	.904	.376	.084	.078	.949	.527	.087	.081	.951
MRNNMI	60%	.100	.060	.051	.930	.374	.100	.092	.938	.526	.103	.095	.945
	70%	.098	.053	.048	.912	.374	.088	.085	.940	.528	.091	.088	.939
	80%	.097	.048	.046	.911	.376	.080	.078	.949	.526	.083	.081	.948
DRNNMI	60%	.099	.059	.051	.919	.375	.098	.092	.937	.526	.102	.095	.945
	70%	.098	.053	.048	.913	.374	.088	.085	.939	.528	.091	.088	.941
	80%	.098	.048	.046	.914	.376	.080	.078	.939	.526	.082	.081	.945
MRPMMI	60%	.100	.061	.051	.929	.375	.102	.093	.940	.526	.105	.097	.942
	70%	.098	.053	.049	.915	.374	.090	.085	.941	.528	.093	.088	.939
	80%	.097	.048	.046	.907	.376	.081	.078	.946	.527	.084	.081	.945
$\text{MICE}_{\text{default}}$	60%	.163	.059	.076	.814	.362	.074	.081	.905	.475	.076	.079	.885
	70%	.136	.054	.063	.897	.369	.073	.077	.924	.496	.075	.076	.923
	80%	.120	.050	.053	.927	.373	.072	.073	.937	.507	.074	.074	.936
MICE_{CM}	60%	.154	.058	.075	.837	.364	.075	.083	.903	.482	.077	.082	.887
	70%	.131	.053	.062	.899	.369	.073	.079	.917	.500	.075	.078	.918
	80%	.117	.049	.053	.925	.373	.072	.074	.935	.510	.074	.075	.937
MICE_{INM}	60%	.151	.058	.067	.876	.359	.074	.078	.913	.491	.077	.080	.908
	70%	.129	.053	.058	.926	.365	.073	.076	.927	.506	.076	.078	.929
	80%	.116	.049	.051	.929	.370	.072	.073	.931	.514	.074	.075	.935
$\text{MICE}_{\text{INNM}}$	60%	.138	.057	.067	.889	.446	.079	.089	.818	.416	.077	.079	.686
	70%	.124	.053	.059	.916	.432	.077	.083	.872	.444	.075	.077	.792
	80%	.115	.049	.053	.920	.417	.075	.077	.909	.468	.074	.075	.872

Note. In Scenario 5 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Table F22.

The Results of the Different Robust MI Methods in Scenario 5 for $n = 100$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.090	-	.037	-	.376	-	.063	-	.534	-	.066	-
	70%	.093	-	.034	-	.376	-	.057	-	.532	-	.060	-
	80%	.095	-	.033	-	.377	-	.054	-	.528	-	.056	-
MRIC$_{R^2_{MF}}$	60%	.091	.043	.039	.884	.376	.074	.067	.948	.533	.077	.070	.945
	70%	.093	.038	.036	.899	.376	.064	.060	.950	.531	.066	.062	.951
	80%	.095	.034	.033	.905	.377	.057	.055	.950	.528	.059	.058	.945
MRIC$_{R^2_{CS}}$	60%	.091	.043	.039	.885	.376	.074	.067	.950	.533	.076	.070	.946
	70%	.093	.038	.036	.898	.375	.064	.060	.952	.532	.067	.062	.951
	80%	.095	.034	.034	.901	.377	.057	.056	.949	.528	.059	.058	.945
MRIC$_{R^2_{MZ}}$	60%	.091	.043	.039	.882	.375	.074	.067	.945	.534	.077	.070	.947
	70%	.093	.038	.036	.889	.376	.064	.060	.949	.531	.066	.062	.946
	80%	.095	.034	.034	.903	.377	.058	.055	.950	.528	.059	.057	.947
MRIC$_{R^2_{N}}$	60%	.091	.043	.039	.887	.376	.075	.067	.947	.533	.077	.070	.945
	70%	.093	.038	.036	.895	.376	.064	.060	.949	.531	.066	.063	.945
	80%	.095	.034	.034	.899	.377	.058	.055	.950	.528	.059	.058	.946
MRIC$_{\text{AIC}}$	60%	.091	.042	.039	.885	.376	.073	.067	.944	.534	.075	.070	.942
	70%	.093	.037	.036	.894	.376	.064	.059	.952	.531	.066	.062	.948
	80%	.095	.034	.034	.904	.377	.057	.055	.951	.528	.059	.057	.951
MRIC$_{\text{HL}}$	60%	.091	.049	.040	.887	.376	.086	.070	.962	.533	.089	.073	.961
	70%	.093	.041	.037	.893	.376	.070	.061	.961	.531	.072	.064	.963
	80%	.095	.035	.034	.902	.377	.060	.056	.955	.528	.062	.058	.953
MRNNMI	60%	.091	.043	.039	.884	.376	.074	.067	.946	.533	.077	.070	.945
	70%	.093	.038	.036	.889	.376	.064	.060	.952	.531	.066	.063	.950
	80%	.095	.034	.033	.907	.377	.058	.055	.950	.528	.059	.057	.949
DRNNMI	60%	.091	.042	.039	.886	.376	.072	.067	.946	.533	.075	.070	.945
	70%	.093	.037	.036	.895	.376	.063	.060	.948	.531	.065	.063	.946
	80%	.095	.034	.033	.902	.377	.057	.055	.950	.528	.059	.058	.948
MRPMMI	60%	.091	.045	.039	.890	.375	.078	.068	.950	.533	.080	.071	.949
	70%	.093	.039	.036	.898	.376	.067	.060	.955	.531	.069	.062	.957
	80%	.095	.035	.034	.905	.377	.059	.056	.951	.528	.061	.058	.949
MICE$_{\text{default}}$	60%	.133	.040	.057	.806	.369	.054	.065	.884	.498	.055	.064	.881
	70%	.118	.037	.045	.887	.374	.053	.059	.911	.508	.054	.058	.915
	80%	.110	.034	.038	.912	.376	.051	.054	.927	.514	.053	.055	.925
MICE$_{\text{CM}}$	60%	.127	.039	.056	.815	.370	.054	.067	.876	.502	.055	.066	.875
	70%	.115	.036	.044	.881	.374	.053	.059	.910	.511	.054	.060	.913
	80%	.109	.034	.038	.911	.376	.051	.055	.932	.516	.053	.055	.928
MICE$_{\text{INM}}$	60%	.126	.039	.052	.848	.364	.054	.063	.885	.510	.055	.064	.890
	70%	.114	.036	.043	.895	.369	.053	.057	.912	.517	.054	.059	.925
	80%	.108	.034	.037	.913	.373	.051	.054	.934	.520	.053	.055	.930
MICE$_{\text{INNM}}$	60%	.123	.039	.050	.862	.458	.057	.066	.676	.420	.055	.059	.522
	70%	.115	.037	.042	.898	.439	.055	.059	.780	.446	.054	.056	.680
	80%	.111	.035	.038	.912	.421	.053	.056	.857	.469	.053	.054	.810

Note. In Scenario 5 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Table F23.
The Results of the Different Robust MI Methods in Scenario 5 for $n = 200$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.090	-	.026	-	.375	-	.044	-	.535	-	.046	-
	70%	.092	-	.024	-	.376	-	.041	-	.532	-	.042	-
	80%	.095	-	.023	-	.376	-	.038	-	.529	-	.040	-
$MRIC_{R^2_{MF}}$	60%	.090	.031	.028	.897	.375	.054	.048	.952	.535	.055	.050	.947
	70%	.093	.027	.026	.903	.375	.046	.043	.956	.532	.048	.044	.953
	80%	.095	.025	.024	.919	.376	.041	.040	.948	.529	.042	.041	.949
$MRIC_{R^2_{CS}}$	60%	.090	.031	.028	.897	.376	.053	.048	.951	.534	.055	.049	.951
	70%	.092	.027	.026	.907	.376	.046	.043	.952	.532	.048	.044	.951
	80%	.095	.024	.024	.913	.376	.041	.040	.946	.529	.042	.041	.946
$MRIC_{R^2_{MZ}}$	60%	.090	.031	.028	.889	.375	.054	.048	.952	.535	.055	.049	.950
	70%	.092	.027	.026	.904	.376	.046	.042	.951	.532	.048	.044	.955
	80%	.095	.024	.024	.909	.376	.041	.039	.955	.529	.042	.041	.952
$MRIC_{R^2_{N}}$	60%	.090	.031	.028	.890	.375	.054	.048	.952	.535	.056	.049	.948
	70%	.092	.027	.025	.906	.375	.046	.043	.953	.532	.048	.044	.950
	80%	.095	.025	.024	.913	.376	.041	.039	.953	.529	.042	.041	.945
$MRIC_{AIC}$	60%	.090	.031	.028	.893	.375	.053	.048	.952	.534	.054	.049	.947
	70%	.092	.027	.025	.904	.376	.046	.043	.957	.532	.047	.044	.951
	80%	.095	.024	.024	.909	.376	.041	.040	.946	.529	.042	.041	.943
$MRIC_{HL}$	60%	.090	.033	.028	.904	.376	.058	.049	.961	.535	.060	.050	.955
	70%	.092	.028	.026	.906	.375	.048	.043	.954	.532	.050	.044	.951
	80%	.095	.025	.024	.919	.376	.042	.040	.950	.529	.043	.041	.950
MRNNMI	60%	.090	.031	.028	.893	.375	.054	.047	.953	.535	.056	.049	.951
	70%	.092	.027	.026	.906	.376	.046	.043	.954	.532	.048	.044	.953
	80%	.094	.024	.024	.910	.376	.041	.040	.949	.529	.042	.041	.947
DRNNMI	60%	.090	.030	.028	.888	.376	.052	.047	.949	.534	.054	.049	.945
	70%	.092	.027	.026	.902	.376	.046	.042	.954	.532	.047	.044	.944
	80%	.095	.024	.024	.914	.376	.041	.040	.951	.529	.042	.041	.949
MRPMMI	60%	.090	.035	.029	.907	.376	.060	.050	.955	.534	.062	.051	.958
	70%	.092	.030	.026	.909	.376	.051	.044	.963	.532	.053	.045	.962
	80%	.094	.026	.024	.916	.376	.043	.040	.956	.529	.045	.041	.954
$MICE_{default}$	60%	.116	.027	.039	.819	.373	.039	.048	.876	.511	.039	.047	.882
	70%	.109	.025	.031	.883	.375	.038	.042	.911	.516	.038	.042	.919
	80%	.105	.024	.027	.913	.375	.037	.038	.936	.520	.037	.039	.933
$MICE_{CM}$	60%	.114	.027	.039	.820	.373	.039	.048	.876	.513	.039	.048	.885
	70%	.108	.025	.031	.877	.375	.038	.043	.913	.517	.038	.042	.918
	80%	.104	.024	.027	.910	.375	.037	.039	.935	.520	.037	.039	.934
$MICE_{INM}$	60%	.113	.026	.037	.832	.366	.038	.047	.868	.521	.039	.047	.895
	70%	.106	.025	.030	.887	.370	.037	.042	.913	.523	.038	.042	.924
	80%	.103	.024	.026	.913	.372	.036	.038	.937	.524	.037	.039	.935
$MICE_{INNM}$	60%	.117	.028	.035	.851	.461	.041	.047	.445	.421	.039	.042	.271
	70%	.112	.026	.030	.900	.440	.039	.042	.621	.447	.038	.039	.486
	80%	.109	.025	.027	.915	.420	.038	.039	.778	.471	.037	.038	.708

Note. In Scenario 5 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Table F24.

The Results of the Different Robust MI Methods in Scenario 5 for $n = 500$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.090	-	.016	-	.375	-	.027	-	.535	-	.029	-
	70%	.092	-	.015	-	.375	-	.026	-	.533	-	.027	-
	80%	.094	-	.014	-	.376	-	.024	-	.530	-	.025	-
$\text{MRIC}_{R^2_{MF}}$	60%	.090	.020	.017	.889	.375	.035	.030	.956	.536	.035	.031	.951
	70%	.092	.018	.016	.899	.375	.030	.027	.957	.533	.030	.028	.947
	80%	.094	.016	.015	.912	.376	.026	.025	.956	.530	.027	.026	.949
$\text{MRIC}_{R^2_{CS}}$	60%	.090	.020	.018	.886	.375	.035	.030	.959	.535	.036	.031	.946
	70%	.092	.018	.016	.901	.375	.030	.027	.956	.533	.030	.028	.949
	80%	.094	.016	.015	.916	.375	.026	.025	.954	.530	.027	.026	.942
$\text{MRIC}_{R^2_{MZ}}$	60%	.090	.020	.018	.883	.375	.035	.030	.962	.536	.036	.031	.947
	70%	.092	.018	.016	.902	.375	.029	.027	.955	.533	.030	.028	.943
	80%	.094	.016	.015	.914	.376	.026	.025	.953	.530	.027	.026	.944
$\text{MRIC}_{R^2_{N}}$	60%	.090	.020	.018	.888	.375	.035	.030	.958	.536	.036	.031	.943
	70%	.092	.018	.016	.902	.375	.030	.027	.959	.533	.031	.028	.948
	80%	.094	.016	.015	.915	.376	.026	.025	.953	.530	.027	.026	.948
MRIC_{AIC}	60%	.090	.020	.018	.884	.375	.034	.030	.953	.535	.035	.031	.944
	70%	.092	.017	.016	.900	.375	.029	.027	.951	.533	.030	.028	.946
	80%	.094	.016	.015	.909	.376	.026	.025	.950	.530	.027	.026	.942
MRIC_{HL}	60%	.090	.021	.018	.892	.375	.036	.030	.961	.535	.037	.031	.953
	70%	.092	.018	.016	.910	.375	.030	.027	.960	.533	.031	.028	.952
	80%	.094	.016	.015	.911	.376	.026	.025	.958	.530	.027	.026	.947
MRNNMI	60%	.090	.020	.018	.889	.375	.035	.030	.967	.535	.036	.031	.950
	70%	.092	.018	.016	.904	.375	.030	.027	.957	.533	.031	.028	.947
	80%	.094	.016	.015	.914	.376	.026	.025	.953	.530	.027	.026	.945
DRNNMI	60%	.090	.020	.018	.884	.375	.034	.030	.956	.535	.035	.031	.944
	70%	.092	.017	.016	.903	.375	.029	.027	.957	.533	.030	.028	.944
	80%	.094	.016	.015	.915	.376	.026	.025	.955	.530	.027	.026	.943
MRPMMI	60%	.090	.027	.019	.909	.375	.046	.032	.976	.535	.047	.034	.970
	70%	.092	.022	.017	.916	.375	.037	.029	.970	.533	.038	.030	.964
	80%	.094	.018	.015	.928	.376	.030	.026	.969	.530	.031	.027	.963
$\text{MICE}_{\text{default}}$	60%	.106	.017	.024	.814	.374	.025	.031	.874	.519	.025	.030	.886
	70%	.104	.016	.020	.878	.375	.024	.027	.911	.521	.024	.027	.917
	80%	.102	.015	.017	.915	.376	.023	.024	.938	.522	.024	.025	.928
MICE_{CM}	60%	.106	.017	.024	.811	.375	.025	.031	.867	.520	.025	.031	.883
	70%	.103	.016	.020	.877	.375	.024	.027	.914	.522	.024	.027	.918
	80%	.102	.015	.017	.909	.376	.023	.024	.935	.522	.024	.025	.930
MICE_{INM}	60%	.103	.016	.023	.829	.368	.024	.031	.863	.528	.025	.030	.881
	70%	.102	.016	.019	.883	.371	.024	.027	.912	.528	.024	.027	.915
	80%	.101	.015	.017	.914	.373	.023	.024	.934	.526	.024	.025	.929
$\text{MICE}_{\text{INNM}}$	60%	.115	.018	.022	.823	.463	.026	.030	.112	.422	.025	.026	.027
	70%	.111	.017	.019	.875	.441	.025	.027	.279	.448	.024	.025	.134
	80%	.108	.016	.017	.917	.421	.024	.025	.547	.472	.024	.024	.418

Note. In Scenario 5 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.

Table F25.
The Results of the Different Robust MI Methods in Scenario 5 for $n = 1000$

		P(Y = 1) = .100				P(Y = 2) = .376				P(Y = 3) = .523			
Method	RR	AEP	ASE	SD	CR	AEP	ASE	SD	CR	AEP	ASE	SD	CR
CCA	60%	.090	-	.012	-	.374	-	.020	-	.536	-	.020	-
	70%	.092	-	.011	-	.375	-	.018	-	.533	-	.019	-
	80%	.095	-	.010	-	.375	-	.017	-	.530	-	.017	-
$\text{MRIC}_{R^2_{MF}}$	60%	.090	.014	.013	.868	.374	.025	.021	.963	.536	.025	.022	.938
	70%	.092	.012	.012	.880	.375	.021	.019	.962	.533	.022	.020	.942
	80%	.095	.011	.011	.905	.375	.018	.017	.957	.530	.019	.018	.945
$\text{MRIC}_{R^2_{CS}}$	60%	.090	.014	.013	.862	.374	.025	.021	.962	.536	.025	.022	.935
	70%	.092	.013	.012	.881	.375	.021	.019	.957	.533	.022	.020	.939
	80%	.095	.011	.011	.909	.375	.019	.018	.959	.530	.019	.018	.942
$\text{MRIC}_{R^2_{MZ}}$	60%	.090	.014	.013	.862	.374	.025	.021	.965	.536	.025	.022	.938
	70%	.092	.012	.012	.878	.375	.021	.019	.957	.533	.022	.020	.932
	80%	.095	.011	.011	.902	.375	.018	.017	.957	.530	.019	.018	.950
$\text{MRIC}_{R^2_{N}}$	60%	.090	.014	.013	.864	.375	.025	.021	.958	.535	.025	.022	.937
	70%	.092	.012	.012	.881	.375	.021	.019	.957	.533	.022	.020	.935
	80%	.095	.011	.011	.901	.375	.018	.017	.960	.530	.019	.018	.945
MRIC_{AIC}	60%	.090	.014	.013	.859	.375	.025	.021	.960	.536	.026	.022	.935
	70%	.092	.013	.012	.884	.375	.021	.019	.960	.533	.022	.020	.938
	80%	.095	.011	.011	.901	.376	.018	.017	.960	.530	.019	.018	.947
MRIC_{HL}	60%	.090	.015	.013	.872	.374	.026	.022	.966	.536	.027	.022	.946
	70%	.092	.013	.012	.886	.375	.022	.019	.966	.533	.023	.020	.943
	80%	.095	.011	.011	.906	.375	.019	.018	.963	.530	.020	.018	.949
MRNNMI	60%	.090	.015	.013	.866	.375	.025	.021	.961	.535	.026	.022	.942
	70%	.092	.013	.012	.882	.375	.021	.019	.959	.533	.022	.020	.941
	80%	.095	.011	.011	.907	.375	.019	.017	.961	.530	.019	.018	.945
DRNNMI	60%	.090	.014	.013	.855	.374	.024	.021	.958	.536	.025	.022	.933
	70%	.092	.012	.012	.878	.375	.021	.019	.958	.533	.022	.020	.936
	80%	.094	.011	.011	.900	.375	.018	.017	.959	.530	.019	.018	.944
MRPMMI	60%	.090	.023	.015	.907	.375	.039	.025	.983	.536	.040	.026	.972
	70%	.092	.018	.013	.911	.375	.031	.021	.982	.533	.032	.022	.973
	80%	.095	.014	.011	.932	.375	.024	.019	.975	.530	.025	.019	.970
$\text{MICE}_{\text{default}}$	60%	.104	.012	.017	.812	.375	.017	.022	.871	.521	.018	.022	.880
	70%	.102	.011	.014	.876	.375	.017	.019	.912	.523	.017	.019	.915
	80%	.101	.011	.012	.922	.376	.016	.017	.939	.523	.017	.017	.941
MICE_{CM}	60%	.103	.012	.017	.811	.375	.017	.022	.873	.522	.018	.022	.880
	70%	.102	.011	.014	.872	.375	.017	.019	.911	.523	.017	.019	.918
	80%	.101	.011	.012	.922	.376	.016	.017	.937	.523	.017	.017	.943
MICE_{INM}	60%	.101	.011	.016	.825	.369	.017	.022	.853	.530	.018	.022	.862
	70%	.100	.011	.014	.870	.371	.017	.019	.903	.529	.017	.019	.907
	80%	.100	.011	.012	.918	.373	.016	.017	.934	.527	.017	.017	.941
$\text{MICE}_{\text{INNM}}$	60%	.114	.012	.016	.750	.464	.018	.021	.009	.422	.018	.019	.000
	70%	.111	.012	.014	.827	.441	.018	.019	.058	.448	.017	.018	.010
	80%	.108	.011	.012	.893	.421	.017	.018	.265	.471	.017	.017	.132

Note. In Scenario 5 one of the outcome models and one of the response models is correctly defined. *CCA* refers to a complete-case analysis in which the units with missing values of Y were removed. *RR* refers to the response rate, *AEP* to the average estimated proportion, *ASE* to the average standard error, *SD* to the standard deviation and *CR* to the coverage rate.