Ludwig-Maximilians-Universität München

Department of Statistics

Various approaches to deal with missing data when estimating causal effects with targeted maximum likelihood estimation

---

Master Thesis

Author: Christoph Wiederkehr

Supervisors: Prof. Dr. Christian Heumann
Prof. Dr. Michael Schomaker

Date: July 11, 2023

Abstract

The integrity of longitudinal studies is deeply interconnected with the accurate management of missing data, a ommnipresent issue that can introduce significant bias if improperly addressed. This paper aims to reproduce and build upon a previous simulation study that evaluated nine methods for handling missing data in the context of estimating the average treatment effect (ATE) using targeted maximum likelihood estimation (TMLE). Building on this, the current study implements and assesses a variety of TMLE methods in R, testing their robustness under extended data-generating processes. Scenarios where the ATE is not recoverable, and complete-case methods yield satisfactory or superior performance are specifically examined. Results align with previous findings that no single method consistently outperforms across all missing directed acyclic graphs (m-DAGs). Non-MI methods like complete-case and extended TMLE exhibit lower relative bias for certain m-DAGs but underestimate model standard errors (ModSEs) especially in complex scenarios. Conversely in the reproduced simulation, MI methods, notably MI Amelia, demonstrate precision and low bias, resulting in low root mean square error (RMSE), even though there is an overestimation of ModSEs. Certain considered TMLE implementations perform superiorly across all scenarios. These findings underscore the necessity of considering the specific scenario and m-DAG when selecting an imputation method, and emphasize the implications of increased complexity on the reliability of imputation and estimation methods.

Acknowledgements

First of all, I want to thank my supervisors, Prof. Dr. Christian Heumann and Prof. Dr. Michael Schomaker, for their willingness to accompany and supervise me throughout this thesis. Their regular support, valuable advice, and guidance have been instrumental in shaping my ideas and overcoming challenging moments.

I would also like to express my gratitude to Prof. Dr. Christian Heumann for granting me access to his LRZ AI account to use the Interactive RStudio Server, which made the simulation study possible.

Furthermore, I am grateful to Prof. Dr. Michael Schomaker for his assistance with the literature research and providing example codes for specific function implementations. His guidance and discussions have been invaluable in expanding my understanding and refining my work.

Contents

1 Introduction

Longitudinal studies are an indispensable tool in epidemiologic research, providing critical insights into health trends, disease processes, and the impact of interventions over time. Central to the utility and validity of these studies is the principle of causal inference. In this domain, targeted maximum likelihood estimation (TMLE) has emerged as a robust method for causal effect estimation. Esteemed for its double-robustness, TMLE offers the potential for unbiased estimates even under conditions where some model assumptions are not fully satisfied. However, the enduring challenge of handling missing data persists in the deployment of TMLE, as it does across the broader landscape of epidemiologic research. Missing data are a common occurrence in longitudinal studies. If not correctly managed, they have the potential to introduce bias, decrease statistical power, and complicate the data analysis process. As such, the manner in which researchers approach missing data is of critical importance. Directed acyclic graphs (DAGs), or causal diagrams, serve as vital conceptual tools in this regard. They are frequently used to visualize the relationships among various variables and anticipate potential bias due to confounding or selection bias. When dealing with missing data, these diagrams can be instrumental in understanding the underlying data-generating mechanism, guiding the choice of the appropriate methodology for handling missing data. The correct strategy can range from simple methods like complete-case analysis to more advanced techniques like multiple imputation, depending on whether the data are missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR). This study was initially motivated by the intent to reproduce and expand on the research of Dashti et al. (2021). The paper brought into focus the performance of complete-case methods in scenarios involving missing directed acyclic graphs (m-DAGs) where the target parameter is whether recoverable and not recoverable. It was found that complete-case methods often provided sufficiently good performance compared to standard alternative methods, such as multiple imputation by chained equations (MICE) The present research aims to build upon this work by conducting a comprehensive simulation study that evaluates various methods for managing missing data within the diverse TMLE implementations available in R (R Core Team, 2023).

2 Motivation/Aim

The initial motivation for this thesis was to reproduce the simulation study of Dashti et al. (2021). In summary, they evaluated the performance of nine available approaches to handle missing data when estimating the average treatment effect (ATE) using targeted maximum likelihood estimation (TMLE) with data-adaptive methods in different complex datasets. This was based on motivating data from the Victorian Adolescent Health Cohort Study. For the simulation of missing data processes, various missing directed acyclic graphs (m-DAGs) were used, encapsulating conventional missing data mechanisms observed in epidemiological research involving incomplete data on exposure, outcome, and confounding variables. Furthermore, the focus should be on m-DAGs where the target parameter (ATE) is not recoverable, and complete-case methods still provide sufficiently good or better performance than the

standard alternative methods (like multiple imputation with multivariate imputation by chained equations). Various TMLE implementations in R (R Core Team, 2023) are also to be used and compared under similar settings and scenarios. In addition, the results and approaches of Dashti et al. (2021) should be tested for their robustness under different extended data-generating processes.

3 Theoretical Background

3.1 Missing Data

In the context of partially observed data, we characterize the missing mechanism by the conditional distribution of missing indicator matrix R given the complete data matrix X. The complete data X is a matrix consisting of n observations and p features, each feature being represented by an index $j = 1, \ldots, p$ and each observation by an index $i = 1, \ldots, n$. Each feature can be thought of as a column in the matrix. Every element of R_j is an indicator of missing data, and it denotes the status of each component of a feature vector X_j. More specifically, if an individual data point in feature X_j, let's denote it as X_{ij}, is observed, then the corresponding entry in R, denoted as $R_{ij} = 1$. Conversely, $R_{ij} = 0$ if X_{ij} is missing.The term 'missing pattern' represents the distribution or propensity of missing data, as indicated by R, across the complete data set X. Analyzing this pattern provides insights into the nature of the missing data within the dataset. The most typical categorization of missing mechanisms is as follows (Little and Rubin, 2002):

- **Missing Completely at Random (MCAR):** This occurs when the missing pattern is completely random and independent of the data. In other words, the probability of a data point being missing does not depend on the data itself: $P(R \mid X) = P(R)$.
- **Missing at Random (MAR):** The missing pattern is independent of unobserved values X_{miss} conditioned on observed values $X_{\text{obs}} : P(R \mid X) = P\left(R \mid X_{\text{obs}}\right)$.
- **Missing Not at Random (MNAR):** The missing pattern may depend on unobserved values. This means that the missingness can depend on the actual values of the data, whether observed or not.

Mechanisms such as MCAR and MAR are sometimes considered ignorable. This implies that these mechanisms may not affect likelihood-based inference and thus can be disregarded (Rubin, 1975). Many imputation methods handle MCAR or MAR data effectively. However, data being MNAR present more significant challenges for imputation methods, necessitating the modeling of the missing mechanism itself. In the context of incomplete data, the MAR assumption is generally considered to enable unbiased estimation when using suitable methods. Nonetheless, the importance of assessing the plausibility of MAR and conducting sensitivity analyses under MNAR scenarios is underscored (van Buuren, 2018).

3.1.1 m-Dags

With multivariable missingness, understanding the concept of MAR can be tricky. Interestingly, under several MNAR scenarios, it is possible to obtain unbiased estimations using methods commonly associated with MAR. As an alternative framework, directed acyclic graphs (DAGs) have been suggested for specifying practically accessible assumptions beyond the MAR-MNAR dichotomy (Karthika and Judea, 2019). Later in the conducted simulation, a general point-exposure study is considered, where (i.e. in reproduced data) incomplete exposure A, incomplete outcome Y, an auxillary variable B, a set of complete confounders W_1 and W_5, and a set of incomplete confounders W_2, W_3 and W_4 are present. These variables can be binary, continuous, or of any other type. Causal relationships between variables are assumed to be depicted by DAGs, which utilize nodes connected by directed arrows. The absence of relationships is encoded by the omission of variables or arrows (Greenland S., 1999). Furthermore U is introduced, representing all completely unmeasured common causes of the exposure and outcome, suggesting that the set of measured covariates is sufficient for confounding adjustment. The relationships between the measured covariates are not depicted. The DAG that would be assumed if the data were complete, named complete-data DAG (c-DAG), can be considered in figure 1. However, a general algorithm for deciding how to manage missing data based on a specific DAG is still lacking. In addressing this, Missingness Directed Acyclic Graphs (m-DAGs) have been formulated (Moreno-Betancur et al., 2018), encapsulating conventional missing data mechanisms observed in epidemiological research involving incomplete data on exposure, outcome, and confounding variables. These m-DAGs include variable-specific missingness indicators to represent assumptions about missingness in each variable. For instance, a missingness indicator for A is defined as $M_A = 1$ if A is missing and $M_A = 0$ if not. A similar definition applies to M_Y. As for the incomplete confounders, they are grouped together, with the missingness indicator $M_W = 1$ if any of the components of W_2, W_3 and W_4 is missing, and $M_W = 0$ otherwise. The basic MCAR or trivial m-DAG, denoted as T, is exemplified in figure 2, with U intentionally omitted for a clearer view. To replicate the exact m-DAGs from Dashti et al. (2021), a slight modification to the fourth assumption from Moreno-Betancur et al. (2018), which states that there are no direct arrows between the missingness indicators, is needed when constructing m-DAGs. This is because the missingness indicators are used to better control joint missing portions in the variables, allowing associations among themselves. Although such causal relationships would be infrequent in the point-exposure study, given that all variables and their missingness indicators, except for the outcome, are measured simultaneously and therefore cannot cause one another. In short, all extensions of m-DAG A (presented in figure 2) were sorted into 16 categories. This classification was based on whether there were arrows from 1) confounders and/or the exposure variable to the missingness indicators of other variables, 2) confounders and/or the exposure variable to their own missingness indicators, 3) the outcome variable to missingness indicators of other variables, and 4) the outcome variable to its own missingness indicator. The m-DAG with the most arrows was chosen as the canonical representative of each class due to its general nature. Out of the 16 resulting canonical m-DAGs, ten and additionally m-Dag T were selected, as illustrated in Figure 2. These were chosen for their representation of all distinct recoverability scenarios and for having the most arrows. Moreover, for each m-DAG, it is crucial

to determine whether common target parameters are recoverable. In cases where these parameters are not recoverable, sensitivity analyses may be required. This work focuses on the recoverability of the conditional distribution of outcome Y, since it is vital for the exposure-outcome association adjusted for confounding through regression (and target parameter ψ_{ATE}). Further formal and technical details including proofs are afforded in the Web Appendix of Moreno-Betancur et al. (2018). By definition, recoverable parameters can be consistently estimated using only the available data, provided an appropriate method is used. Available-case analysis involves estimating the target parameter using only records with complete data on the involved variables. Importantly, if a parameter is recoverable in a specific canonical m-DAG, then it is recoverable in all m-DAGs within the class that it symbolizes. This is because these m-DAGs are derived by merely removing arrows. In brief, when the outcome variable variable causes its own missingness, the conditional distribution of the outcome is not recoverable (m-Dags G, H, J). In addition to this, the conditional distribution in m-Dags F, I is not recoverable because of the collider structures $A \rightarrow M_Y \leftarrow Y$ and $W_2, W_3, W_4 \rightarrow M_W \leftarrow Y$. These structures preclude identification of this parameter. For the m-Dag C, the joint distribution can be expressed in terms of available data. This generally makes the conditional distribution of Y recoverable. Though, it's not expressible using an available-case approach. It's necessary to note that modifying the fourth assumption may result not just in an increase in the number of scenarios but also require further theoretical work. Nonetheless the conditional independence attributes for the respective m-DAGs remain the same in m-DAGs A, B, D, E, and T where the conditional distribution of Y can be represented as the conditional distribution among the complete cases. Thus providing theoretical recoverability for the exposure-outcome association. However, for other recoverable parameters, potential selection bias could be introduced by available-case analysis. Methods like multiple imputation (MI) can possibly circumvent the selection bias in some situations with available-case analysis. Yet, these methods bring into play parametric assumptions that extend beyond those of the analysis model (i.e., the outcome regression). This could potentially lead to a precision gain in comparison to available case analysis, but it might also induce misspecification bias (Moreno-Betancur et al., 2018). The unbiased nature of multiple imputation is largely contingent on the quality of the parametric assumptions concerning the target parameter, which connects to the concept of 'congeniality' between the imputation and analysis models (White et al., 2010).

3.1.2 MI - with mice and Amelia as R implementations

Missing data is a common issue in statistical analysis and machine learning, presenting challenges to making valid inferences from the data. Multiple Imputation (MI), a technique developed by Rubin (1977), has become a widely accepted approach for handling missing data. The power of MI lies in its ability to generate multiple imputations, which captures together with incorporated random components the uncertainty due to missingness (White et al., 2010). MI replaces each missing data point by generating m independent simulated sets of values, drawn from the posterior predictive distribution of the missing data conditional on the observed data. This differs from single imputation methods, which

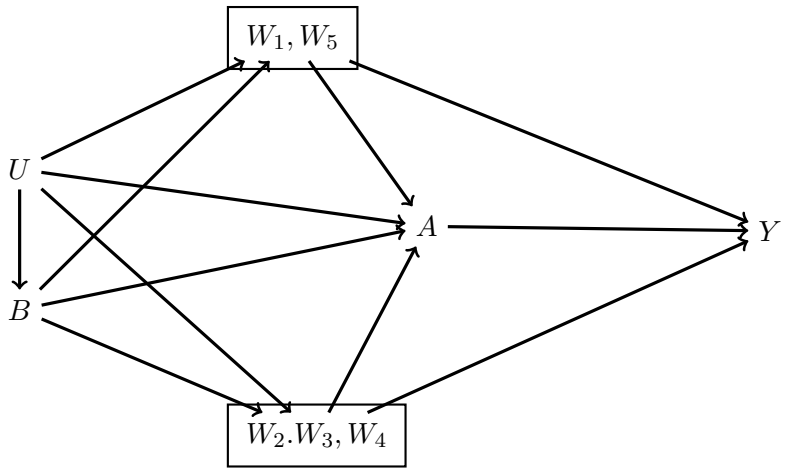


Figure 1: Complete-data directed acyclic graph (c-DAG) for a general point-exposure study. Exposure A, outcome Y, auxillary variable B, a set of complete confounders W_1 and W_5, set of incomplete confounders W_2, W_3 and W_4 and U as unmeasured common causes of the exposure and outcome. Figure has been adapted from Moreno-Betancur et al. (2018).

replace missing values once, often leading to underestimated variance and overestimated significance (van Buuren, 2018). The three stages of MI include generating multiply imputed datasets, analyzing these datasets, and combining the estimates from these datasets (White et al., 2010):

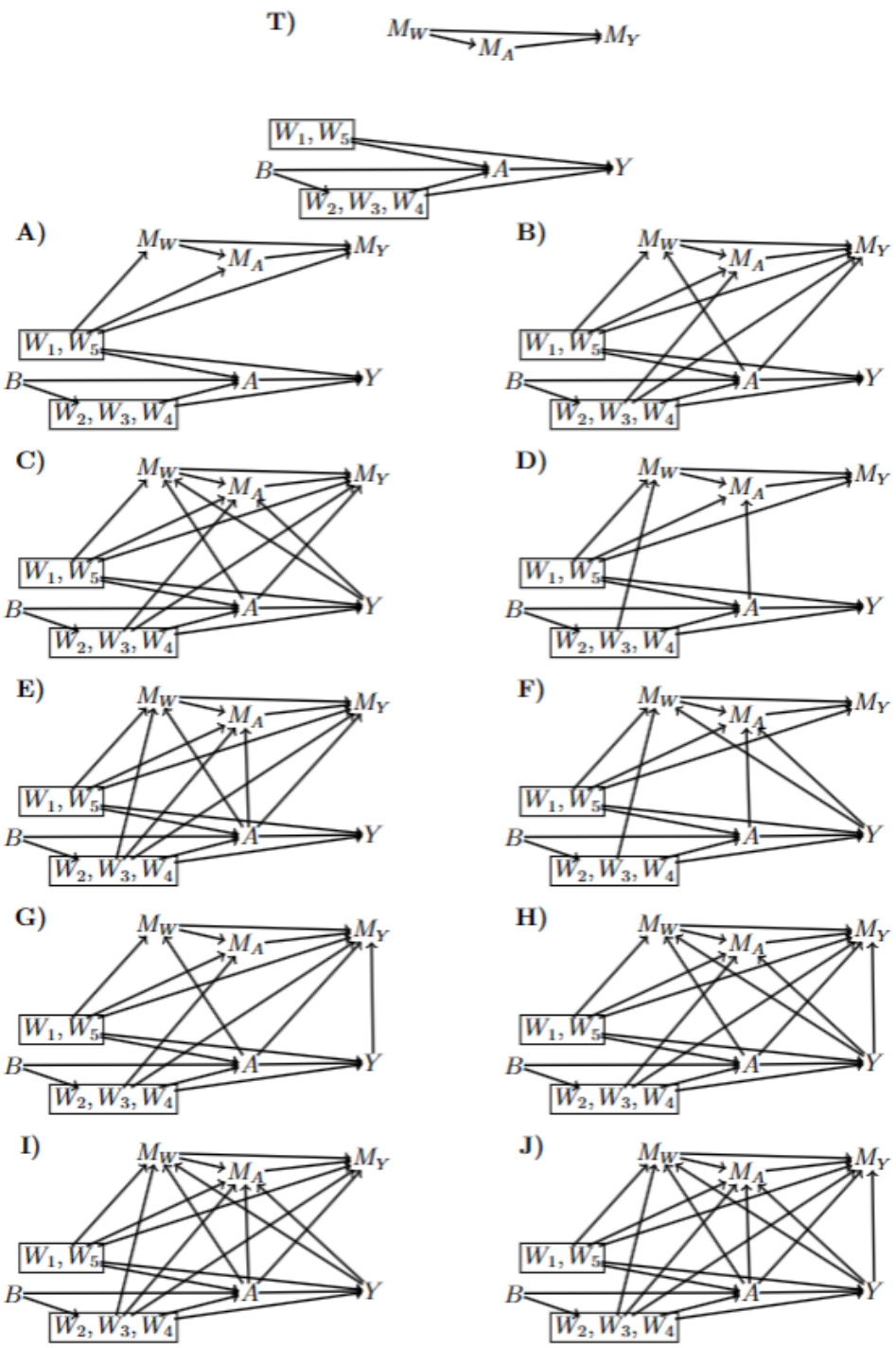
- The initial stage involves creating m complete datasets. For an incomplete variable Z, this is achieved by creating an imputation model that regresses Z on a set of variables, denoted as $X_1, X_2, \ldots, X_p$, which have complete data among individuals with observed Z values. If $\hat{\theta}$ and V are the set of estimated regression parameters and their corresponding covariance matrix from fitting the imputation model, the process proceeds as follows:

  1. A random draw, θ^*, is obtained from the posterior distribution, which is often approximated by $\theta^* \sim \mathcal{N}_p(\hat{\theta}, V)$.
  2. Imputations for Z are drawn from the posterior predictive distribution of Z using θ^* and the appropriate probability distribution.

- Once the process of multiple imputation is completed, each dataset that has been imputed is used to separate analysis utilizing standard methods for complete data. Given that the missing values have been substituted with varying imputations, the outcomes from these m analyses will not be identical.

- The m estimates derived from each respective analysis are subsequently integrated into a comprehensive estimate and a variance-covariance matrix using Rubin's rules (Rubin, 1987). The combined variance-covariance matrix encapsulates both within-imputation variability and between-imputation variability. Assume $\hat{\psi}_l$ is an estimate of a univariate or multivariate parameter of interest procured from the lth imputed dataset, and WV_l is the estimated variance of $\hat{\psi}_l$. The

Figure 2: Missingness directed acyclic graphs (m-DAGs) illustrating the missingness scenarios considered in the simulation study. Figure has been adapted from Dashti et al. (2021).



For simplicity of exposition, confounders without missing data (W_1 and W_5) are presented on a single node and confounders with missing data (W_2, W_3, W_4) on another single node. Also, only one missingness indicator has been included for confounders with missing data (M_W), coded as 1 when any of the variables W_2, W_3, W_4 have missing data and as 0 when none has missing data. The m-DAG T (trivial m-DAG) represents the simplest missingness scenario and corresponds to a missing completely at random mechanism.

combined estimate $\hat{\psi}$ is the mean of the individual estimates:

$$\hat{\psi} = \frac{1}{m} \sum_{l=1}^{m} \hat{\psi}_l.$$

The total variance of $\hat{\psi}$ is computed from the within-imputation variance $WV = (1/m) \sum_{l=1}^{m} WV_l$ and the between-imputation variance $BV = \frac{1}{m-1} \sum_{l=1}^{m} \left(\hat{\psi}_l - \hat{\psi}\right)^2$:

$$\text{Var}(\hat{\psi}) = WV + \left(1 + \frac{1}{m}\right) BV$$

(van Buuren, 2018).

There are two general approaches that have emerged for imputing multivariate data: joint modeling (JM) and fully conditional specification (FCS). The latter is also known as multivariate imputation by chained equations (MICE).

The **Amelia** R-package (Honaker et al., 2011) offers a unique approach to missing data imputation by assuming the hypothetically complete data X follows a multivariate normal distribution thus using the JM approach. While the multivariate normal distribution may appear to be a rudimentary approximation of the true data distribution, evidence suggests that this model holds up comparably well to more complex models, even when dealing with categorical or mixed data. Furthermore, transformations of many types of variables can often make this normality assumption more plausible. This approach is expressed as:

$$X \sim \mathcal{N}_p(\mu, \Sigma)$$

which implies that X, the dataset, has a multivariate normal distribution with parameters $\theta = (\mu, \Sigma)$. **Amelia** implies that the missing data pattern is dependent only on the observed data X_{obs} and not on the unobserved data X_{mis}. This is mathematically represented as before:

$$P(R \mid X) = P\left(R \mid X_{\text{obs}}\right).$$

After deriving the likelihood of the observed data and applying the law of iterated expectations, the posterior is found:

$$P\left(\theta \mid X_{\text{obs}}\right) \propto P\left(X_{\text{obs}} \mid \theta\right) = \int p(X \mid \theta)\, dX_{\text{mis}}.$$

Amelia employs the expectation-maximization with bootstrapping (EMB) algorithm, combining the classic expectation-maximization (EM) algorithm with a bootstrap approach, to draw from the posterior. In the bootstrap step, the observed data are resampled with replacement, and the EM algorithm

is applied to the resampled data to obtain bootstrap estimates of the parameters. These bootstrap estimates are used to approximate the posterior distribution of the parameters given the observed data. With the posterior of the complete-data parameters, imputations are created by drawing values of X_{mis} from its distribution conditional on X_{obs} and the draws of θ.

The **mice** R-package (van Buuren and Groothuis-Oudshoorn, 2011), which implements the MICE approach, is a practical tool for handling missing values. MICE works also under the MAR assumption. It iteratively imputes missing data, generating a series of plausible imputed datasets. Let the hypothetically complete data X be a partially observed random sample from the p variate multivariate distribution $P(X \mid \theta)$. It is assumed that the multivariate distribution of X is completely specified by θ, a vector of unknown parameters. Instead of trying to estimate the joint distribution of all the variables, MICE handles each variable individually, allowing the model to accommodate variables of different types and distributions. Starting from a simple draw from observed marginal distributions, the general structure of the MICE algorithm involves iteratively sampling from the conditional distribution of each variable with missing data, given the current imputed values of all other variables:

$$\begin{aligned} &P\left(X_1 \mid X_{-1}, \theta_1\right) \\ &\vdots \\ &P\left(X_p \mid X_{-p}, \theta_p\right). \end{aligned}$$

The parameters $\theta_1, \ldots, \theta_p$ are specific to the respective conditional densities and are not necessarily the product of a factorization of the true joint distribution $P(X \mid \theta)$. The main concern of MICE stems from the fact that the conditionally specified models may be incompatible in the sense that the joint distribution cannot exist. Yet, simulation studies suggest that the consequences of incompatibility on the quality of imputations are usually not serious in practice (van Buuren and Groothuis-Oudshoorn, 2011).

3.2 Super Learner

A Super Learner is an algorithm that utilizes an ensemble of diverse machine learning algorithms for outcome prediction. The optimal combination of included models is selected by the Super Learner, thereby producing an output that is expected to perform at least as well as the best individual predictor included in the ensemble. The theoretical underpinning of Super Learning revolves around stacking, an ensemble learning technique that combines multiple predictive models to generate a new model. Super Learner differs from conventional stacking techniques by dynamically assigning weights to each base learner based on their performance (van der Laan et al., 2007). The Super Learner algorithm operates through a series of steps:

1. The dataset is split into training and validation sets.

2. Each candidate learning algorithm is fitted on the training set.

3. Each of these algorithms is used to predict outcomes on the validation set.

4. These predictions are used as inputs to another learning algorithm (the meta-learner) that predicts the outcome.

What makes a Super Learner powerful is the final step. Instead of using a simple technique like majority voting or averaging to combine the predictions, a Super Learner uses another learning algorithm to determine the optimal combination of the base learner's predictions (Naimi and Balzer, 2018). More theoretical: Given a set of k learning algorithms, denoted as $\{g_1, g_2, \ldots, g_k\}$ and later called 'library'. The objective of the Super Learner, denoted S, is to find the best combination of these candidate algorithms, defined as follows:

$$S = \text{argmin}[L(Y, f(X))],$$

where Y represents the dependent variable, X the independent covariates and $f(X)$ is a weighted combination of $\{g_1, g_2, \ldots, g_k\}$, i.e., $f(X) = w_1 * g_1(X) + w_2 * g_2(X) + \ldots + w_k * g_k(X)$. $L(Y, f(X))$ represents the loss function, a quantification of the difference between the predicted and the actual outcome and argmin $[\ L(Y, f(X))]$ implies the selection of $f(X)$ that minimizes the loss function. The Super Learner algorithm optimizes the weights $w_1, w_2, \ldots, w_k$ such that the loss function is minimized (van der Laan et al., 2007). Distinct from traditional Super Learner models that generally employ logistic regression or other generalized linear models as a meta-learner, non-negative least squares (NNLS) is utilized as a meta-learner within the Super Learner framework in the implementation by Polley et al. (2021), which is later used in the simulation study. The use of NNLS as a meta-learner ensures that the weights derived by the Super Learner algorithm are non-negative and sum up to 1 (Naimi and Balzer, 2018). This approach provides interpretability to the model, as negative weights could imply inverse relationships. Furthermore, if a learning algorithm (i.e. g_2) receives a weight of $w_2 = 0$, it simply indicates that g_2 doesn't contribute to enhancing the predictive accuracy. Some benefits of Super Learner over parametric specifications are (van der Laan and Rose, 2011):

1. **Reduction in model misspecification risks:** Parametric modeling, reliant on an a priori structure, is susceptible to misspecification if the chosen function doesn't fit the data generation process. By employing an ensemble of models, the Super Learner mitigates this risk, accommodating a broader range of functional forms.

2. **Elimination of data distribution assumptions:** Traditional models often assume a specific data distribution, such as normality, which if violated can compromise prediction accuracy. The Super Learner, incorporating non-parametric methods, is capable of dealing with complex relationships without these assumptions.

3. **Accommodation of high-dimensional data:** High-dimensional data can challenge parametric models due to concerns like multicollinearity. The Super Learner's capacity to handle such data

structures effectively is owed to the inclusion of base learners designed for this task, like tree-based models or support vector machines.

4. **Capture of interactions and non-linearities:** Interaction effects or non-linear relationships, often overlooked by parametric models unless specifically modeled, can be implicitly captured by the Super Learner through its base learners.

Despite these significant advantages, the Super Learner also presents challenges, primarily being computationally intensive and necessitating rigorous validation to avoid overfitting. Nonetheless, when deployed appropriately, it outperforms traditional parametric models in prediction tasks (Balzer and Westling, 2021). As stated in the next Chapter 3.3.2, targeted likelihood estimation (TMLE) is double-robust, meaning it can provide unbiased estimates if either the outcome model or the exposure model is correctly specified, but not necessarily both. The use of Super Learner further enhances this property by increasing the chances that at least one of these models is correctly specified (van der Laan and Rose, 2011).

3.3 Causal inference

Causal inference is the science of determining cause and effect relationships from data. It explores how manipulating a variable affects another, beyond mere correlation. These effects should be studied through experiments that randomly assign individuals to treatment or control groups, such that comparable groups are compared under competing treatments. However, many such experiments, specifically those which involve humans, are either infeasible or unethical (Kurz, 2021). Therefore traditional observational studies often need a statistical adjustment for confounders (variables associated with both the exposure and the outcome) in order to obtain unbiased exposure effect estimates. For the estimation of causal effects, numerous estimators can be used: G-computation methods, propensity score methods or double-robust methods (Hernán and Robins, 2020). This chapter primarily focuses on causal inference using targeted maximum likelihood estimation TMLE (van der Laan and Rose, 2011). Nevertheless, the causal inference framework and its necessary assumptions regarding the average treatment effect (ATE) should be explained first.

3.3.1 Causal inference framework

Causal inference based on the Neyman-Rubin potential outcome framework allows researchers to adjust for confounders under structural causal assumptions (Rubin, 2005). Let A denote a binary exposure, W a vector of potential confounders and Y a continuous outcome. In this setting each individual has a pair of potential outcomes given the binary exposure:

$$Y_i^{(A_i)} = \begin{cases} Y_i^{(1)} & \text{if } A_i = 1 \\ Y_i^{(0)} & \text{if } A_i = 0 \end{cases}$$

with individuals indexed by $i = 1, ..., n$. The outcome when receiving a treatment, is denoted as $Y^{(1)}$, and the outcome when not receiving a treatment, is denoted as $Y^{(0)}$. These quantities are referred to as 'potential' outcomes because it is only possible to observe a single realization of the outcome and not both for an individual. This is also called the 'fundamental problem of causal inference' (Holland, 1986). The causal effect of the treatment can be represented by the difference in potential outcomes, $Y^{(1)}$ and $Y^{(0)}$. It is assumed that the actual observed outcome for individual i is connected to the potential outcomes through $Y_i = Y_i^{(1)} A_i + Y_i^{(0)} (1 - A_i)$. A common causal estimand for aggregated causal effects is the average treatment effect (ATE), defined as $\psi_{ATE} = E\left[Y^{(1)} - Y^{(0)}\right]$ (Kurz, 2021) or rather the statistical target parameter:

$$\begin{aligned}\psi_{ATE} &= E_W\left[E(Y^{(1)} \mid W) - E(Y^{(0)} \mid W)\right] \\ &= \sum_w \left[\sum_y \mathrm{P}(Y = y \mid A = 1, W = w) - \sum_y \mathrm{P}(Y = y \mid A = 0, W = w)\right] \mathrm{P}(W = w)\end{aligned}$$

representing the marginal difference in the outcome Y between the exposed and the unexposed, adjusted for measured confounders W, where the outer expectation averages over the distribution of W (Schuler and Rose, 2016). In order to have a causal interpretation for the ATE several key assumptions are required. In specific, we assume that the stable unit treatment value assumption (SUTVA) (Holland, 1986) is satisfied:

- **Noninterference:** The treatment status of a given individual does not affect the potential outcomes of any other individual.
- **Consistency:** The potential outcome corresponds with the observed treatment: $Y = Y^{(1)} A + Y^{(0)} (1 - A)$.
- **Positivity:** Within strata of W, every individual has a nonzero probability of receiving either exposure condition. This is formalized as $0 < P(A = 1 \mid W) < 1$ for a binary exposure.
- **Conditional exchangeability:** The treatment is independent of the potential outcomes after conditioning on W ($\left\{Y^{(1)}, Y^{(0)}\right\} \perp A \mid W$). Or rather the outcome for those treated would have been the same as for those untreated if untreated subjects had received the treatment. This assume that all confounders have been measured.

(Luque-Fernandez et al., 2018).

3.3.2 TMLE

Targeted maximum likelihood estimation, an efficient and double-robust substitution estimators, is a maximum likelihood based G-computation estimator, first introduced by van der Laan and Rubin (2006). In short, for the ATE, the TMLE procedure requires initial estimates of $E(\mathrm{Y} \mid \mathrm{A}, W)$ and

$P(\mathrm{A} = 1 \mid \mathrm{W})$. A subsequent 'targeting' step is then incorporated that optimises the bias-variance tradeoff for the specific parameter in question, such as the ATE. In the subsequent chapters, the ATE is assumed as the target parameter. Additionally, the use of ensemble and machine-learning algorithms (i.e. Super Learner) to estimate $E(\mathrm{Y} \mid \mathrm{A}, W)$ and $P(\mathrm{A} = 1 \mid W)$ can easily be adopted to avoid model misspecification (Luque-Fernandez et al., 2018).
In that sense non-parametric structural equation modeling (NPSEM) is another concept that will be briefly introduced here, as it offers an alternative framework for defining causal effect parameters (Pearl, 2010) and moreover is compatible with ensemble methods such as Super Learner. NPSEM encapsulates our understanding of the structural causal model through a system of equations, as follows:

$$\begin{aligned} W &= f_W\left(U_W\right), \\ A &= f_A\left(W, U_A\right), \\ Y &= f_Y\left(W, A, U_Y\right), \end{aligned}$$

where U_W, U_A, and U_Y denote exogenous error terms, while W , A, and Y represent, in chronological order, the confounders, the binary treatment of interest, and the outcome, respectively. This NPSEM allows the definition of counterfactual outcomes $Y^{(1)} = f_Y\left(W, A = 1, U_Y\right)$ and $Y^{(0)} = f_Y\left(W, A = 0, U_Y\right)$ and thereby the causal quantity of interest. This general formulation leaves the functions f_W, f_A, f_Y open-ended, enabling more flexibility. It can accommodate exclusion restriction assumptions that enhance identifiability by constraining the range of probability distributions under consideration. Furthermore, it allows for the assumption of parametric forms, thereby providing a versatile foundation for various models. In the context of NPSEM, the randomization assumption is associated with assuming that U_A and U_Y are conditionally independent, given W and with respect to the distribution of potential outcomes $Y^{(1)}$ and $Y^{(0)}$. The NPSEM methodology and the counterfactual framework, although presenting unique constructs for understanding causality, both provide the basis for defining causal effects as parameters within statistical distributions (Gruber and van der Laan, 2012).
The TMLE methodology can be described in the following way: An orthogonal factorization of the likelihood of the data $O = (W, A, Y) \sim \mathrm{P}_0(W)$ is given by

$$\mathcal{L}(O) = \mathrm{P}(Y \mid A, W)\mathrm{P}(A \mid W)\mathrm{P}(W).$$

$\mathrm{P}(W)$ and $\mathrm{P}(Y \mid A, W)$ are refered as the Q portion of the likelihood, $Q = (Q_W, Q_Y)$, and $\mathrm{P}(A \mid W)$ as the g portion of the likelihood. P_0 is an unknown underlying probability distribution. Further define

$$\begin{aligned} \bar{Q}_0(A, W) &= \mathrm{E}(Y \mid A, W), \\ g_0(1 \mid W) &= \mathrm{P}_0(A = 1 \mid W). \end{aligned}$$

The subscript '0' represents the truth, while a subscript 'n' signifies the corresponding quantity derived from the data. The empirical distribution of W is used to estimate $\mathrm{P}_0(W)$, the non-parametric maximum likelihood estimate. The term $\bar{Q}_n(A, W)$ can be obtained by regressing Y on A and W or through the use of ensemble and machine-learning algorithms such as Super Learner. Similarly, the propensity score $g_n(1 \mid W)$ can be obtained this way (Gruber and van der Laan, 2012). With the ATE as the target parameter, the Super Learner substitution estimator is :

$$\hat{\psi}_{MLE_n} = \frac{1}{n} \sum_{i=1}^{n} \left[\bar{Q}_n^0 \left(1, W_i\right) - \bar{Q}_n^0 \left(0, W_i\right) \right].$$

(van der Laan and Rose, 2011), when $A = 1$ (represented as $\bar{Q}_n^0(1, W)$) and $A = 0$ (represented as $\bar{Q}_n^0(0, W)$) for all subjects. Thereby $\bar{Q}_n^0(A, W)$ is the initial estimate of $\bar{Q}_0(A, W)$. The superscripts '0' and subsequently '1' refer to the first and to the updated estimates, respectively. The next step is to update the estimator above toward the parameter of interest. The process of targeting utilizes $g_n(A \mid W)$ to define a one-dimensional model for fluctuating the initial estimator, a strategy often referred to as employing a 'clever covariate'. This clever covariate is defined as:

$$H_n^*(A, W) = \left(\frac{I(A = 1)}{g_n(1 \mid W)} - \frac{I(A = 0)}{g_n(0 \mid W)} \right).$$

A straightforward logistic regression, involving just a single variable (the clever covariate), is subsequently executed for the outcome Y. In the context of this study, Y is continuous and consequently needs to be rescaled to the interval $[0, 1]$. Despite that, this is typically done prior to the initial estimation of $\bar{Q}_0(A, W)$ (van der Laan and Rose, 2011). It is also possible to use the clever covaroate as weight. The offset for this regression is given by logit $\left(\bar{Q}_n^0(A, W)\right)$ and the resulting estimate of the fluctuation parameter ϵ is employed to update the initial estimate $\bar{Q}_n^0(A \mid W)$ yielding a new estimate $\bar{Q}_n^1(A \mid W)$ in the following manner:

$$\text{logit} \left(\bar{Q}_n^1(A, W)\right) = \text{logit} \left(\bar{Q}_n^0(A, W)\right) + \epsilon_n H_n^*(A, W)$$

where ϵ_n is the estimate of ϵ.
The updated fit is then employed to compute the expected outcomes when $A = 1$ (represented as $\bar{Q}_n^1(1, W)$) and $A = 0$ (represented as $\bar{Q}_n^1(0, W)$) for all subjects. These estimates are inserted into the subsequent equation to derive the final TMLE estimation of the ATE:

$$\hat{\psi}_{TMLE_n} = \frac{1}{n} \sum_{i=1}^{n} \left[\bar{Q}_n^1 \left(1, W_i\right) - \bar{Q}_n^1 \left(0, W_i\right) \right]$$

(Li et al., 2022). At its core, TMLE, revolves around the concept of adjusting an initial estimation of the conditional mean outcome, and minimizing a loss function to select the magnitude of the fluctuation. TMLE corresponds to the selection of the negative log-likelihood loss function. Given that TMLE provides a solution for the estimating equation of the efficient influence curve, and this curve fulfills a property known as double robustness, it is guaranteed that TMLE will be asymptotically unbiased provided that either the outcome mechanism $\bar{Q}_0(A, W)$ or the treatment mechanism $g_0(1 \mid W)$ is estimated consistently. When both are consistently estimated, TMLE achieves the semi-parametric efficiency bound, under appropriate regularity conditions (van der Laan and Rubin, 2006). For valid statistical inference, regularity conditions first require that the outcome regression and propensity score estimators converge to their targets at sufficiently fast rates. Secondly, these estimators should not be excessively adaptive, a requirement often referred to as the 'Donsker condition' (Balzer and Westling, 2021). Unfortunatley it is possible to violate this assumptions by an overfit of the models of the data-generating distribution (DGD). For instance by using a Super Learner library with random forests and extreme gradient boosting, and this can occur even when cross-validation is used to choose the resulting fits (Li et al., 2022).
In summary TMLE has several theoretical properties that makes it an attractive method for estimating causal effects from observational data. Competing estimators, falling into the broad classes of maximum likelihood estimation (MLE) and estimating equation methodology, do not have all of its properties and will underperform in many scenarios in comparison to TMLE (van der Laan and Rose, 2011).

3.3.3 CV-TMLE

Despite TMLE being a doubly robust and efficient estimator, its performance can suffer when the initial estimator is overly adaptive (van der Laan and Rose, 2011). Intuitively, if the initial estimator of $\bar{Q}_0(A, W)$ is prone to overfit, there is no realistic residual variation left for the targeting step, making the update incapable to mitigate residual bias. To overcome these drawbacks of TMLE, a modified version called cross-validated targeted maximum likelihood estimation (CV-TMLE) was developed (van der Laan and Zheng, 2010). CV-TMLE incorporates an additional layer of cross-validation for the initial estimator, enhancing the robustness of TMLE in its bias reduction step. As a result, it allows for greater flexibility in using adaptive methods to estimate components of the DGD, while maintaining realistic residual variation in the validation sample (Li et al., 2022). A general CV-TMLE procedure can be described in the following way:

1. Split the data into V independent folds (i.e., keeping track of repeated measures or dependencies is necessary to ensure independence of the V folds). Each fold serves as the validation set, with the remaining data forming the training set. This gives us v-specific sample splits across the validation and training sets.

2. For $v = 1, \ldots, V$, carry out the following steps:

(a) Estimate the propensity score and conditional expectation of the outcome on the training set.

(b) Using the models fitted in the previous step, generate predictions for both the observed treatment assignment and the outcome under the treatment assignment for observations within the validation set. Denote the corresponding estimates as $g_{n,v}(A \mid W)$, and $\bar{Q}^0_{n,v}(A, W)$.

(c) Update $\bar{Q}^0_{n,v}(A, W)$ generated in the previous step by fitting an intercept model (as described in the TMLE updating procedure in the previous subsection 3.3.2) on observations in the validation set. Call the updated fit $\bar{Q}^1_{n,v}(A, W)$.

(d) Generate validation set-specific targeted estimates of the conditional mean outcome under $A = 1$ and $A = 0$ on data in the validation set. Denote the updated estimates $\bar{Q}^1_{n,v}(1, W)$ and $\bar{Q}^1_{n,v}(0, W)$.

(e) Define the v^{th} validation set-specific estimate of the marginal difference in the outcome Y between the exposed and the unexposed, adjusted for measured confounders W, as:

$$\hat{\psi}_{TMLEn,v} = \frac{1}{n_v} \sum_{i \in \text{Val}(v)} \left[\bar{Q}^1_{n,v}(1, W_i) - \bar{Q}^1_{n,v}(0, W_i) \right]$$

where n_v denotes the number of individuals in the validation set v and $\text{Val}(v)$ is the indices i for which O_i is in the validation set.

3. Average over all validation folds to obtain the CV-TMLE, i.e., the estimated marginal difference in the outcome Y between the exposed and the unexposed of the sample-split-specific estimates:

$$\hat{\psi}_{CV-TMLEn,v} = \frac{1}{V} \sum_{v=1}^{V} \hat{\psi}_{TMLEn,v}$$

(Montoya et al., 2021).

3.3.4 Inference

Targeted maximum likelihood estimation constructs estimators based on the efficient IC, which can be used to obtain standard errors. The estimator $\hat{\psi}_{TMLE_n}$ is asymptotically linear and consistent for its true value ψ_{TMLE_0} if it can be expressed as:

$$\hat{\psi}_{TMLE_n} - \psi_{TMLE_0} = \frac{1}{\text{n}} \sum_{\text{i}=1}^{\text{n}} \text{IC}_\text{i} - \text{O}_\text{p}\left(\frac{1}{\sqrt{\text{n}}}\right).$$

By the weak law of the large numbers, the O_p term in equation converges to 0 at a rate of $\frac{1}{\sqrt{n}}$ as the sample size n approaches to infinity. The Influence Curve (IC) is a function of the data and the components that generate data which can be derived for a specified model and target parameter that has a mean of 0 and a finite variance. The central limit theorem is applicable such that in large samples, the variance of the estimator is the variance of the IC divided by n (Luque-Fernandez et al., 2018). The estimator $\hat{\psi}_{TMLE_n}$ possesses thus several noteworthy properties: Firstly, as the sample size increases, its bias approaches 0 at a rate faster than $\frac{1}{\sqrt{n}}$. Secondly when n, is large, its distribution approximates a normal distribution, represented as $n^{1/2}\left(\hat{\psi}_{TMLE_n} - \psi_{TMLE_0}\right) \xrightarrow{d} N\left(0, \sigma_0^2\right)$. This property allows the construction of Wald-type confidence intervals using an estimate of σ_0^2. Lastly, the asymptotic variance of $n^{1/2}\left(\hat{\psi}_{TMLE_n} - \psi_{TMLE_0}\right)$ can be accurately approximated by the sample variance of its estimated influence curve $\widehat{IC}_n$. This is equivalent to $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} \widehat{IC}_n^2\left(O_i\right)$, given that the mean of an influence curve is 0 (Montoya et al., 2021). Even though a wide variety of influence functions and corresponding estimators can be used for a specific target parameter, it is always possible to identify an 'efficient' influence curve. This efficient IC achieves the minimum asymptotic variance under the given set of modeling assumptions. The efficient IC can be estimated as:

$$\widehat{EIC}_{TMLE_n} = \left(Y - \bar{Q}_n^1(A,W)\right)\left[\frac{A}{g_n(W)} - \frac{1-A}{1-g_n(W)}\right] + \left(\bar{Q}_n^1(1,W) - \bar{Q}_n^1(0,W)\right) - \hat{\psi}_{TMLE_n}.$$

The standard error estimate for TMLE can be constructed by multiplying $1/\sqrt{n}$ by the standard deviation of the plug-in efficient influence curve:

$$\hat{\sigma}_{TMLE_n} = \sqrt{\frac{\widehat{\mathrm{Var}}\left(\widehat{EIC}_{TMLE_n}\right)}{n}},$$

where $\widehat{\mathrm{Var}}\left(\widehat{EIC}_n\right)$ represents the 'sample variance' of the estimated IC (Luque-Fernandez et al., 2018). Utilizing the TMLE estimator, it is possible to underestimate the variance of the estimator if $\bar{Q}_0(A,W)$ is estimated data-adaptively on the same data used for evaluating the sample variance of the estimated influence curve. Overfitting, a common issue that arises from using the data twice (for both estimation and evaluation), can be mitigated by employing CV-TMLE confidence intervals. This approach involves sample splitting and helps protect the accuracy of the results. The fold-specific estimate of the working influence curve for CV-TMLE is calculated by estimating, $\bar{Q}_0(A,W)$, and $g_0(A \mid W)$ on the v^{th} . This is then evaluated on the complementary validation sample as follows:

$$\begin{aligned}\widehat{EIC}_{CV-TMLEn,v} = &\left(Y - \bar{Q}_{n,v}^1(A,W)\right)\left[\frac{A}{g_{n,v}(W)} - \frac{1-A}{1-g_{n,v}(W)}\right] + \left(\bar{Q}_{n,v}^1(1,W) - \bar{Q}_{n,v}^1(0,W)\right) \\ &- \hat{\psi}_{CV-TMLEn,v}.\end{aligned}$$

Subsequently, the fold-specific estimate of the variance for the fold-specific estimator can be formulated as:

$$\hat{\sigma}_{CV-TMLEn,v} = \sqrt{\frac{\widehat{\text{Var}}\left(\widehat{EIC}_{CV-TMLEn,v}\right)}{n_v}}.$$

Therefore, the asymptotic variance of CV-TMLE can be conservatively estimated by following the approach mentioned in (Montoya et al., 2021):

$$\hat{\sigma}_{n,CV-TMLE} = \frac{1}{V}\sum_{v=1}^{V}\sigma_{CV-TMLEn,v}.$$

4 Methodology

To evaluate the performance of various methods for handling missing data and different TMLE implementations in *R*, a simulation study was conducted, based heavily on the simulations presented in the work of Dashti et al. (2021). However only 1000 datasets, each with 2000 observations, were generated. The reduction in the number of datasets is due to a faster runtime and also computational constraints. The number of observations remains the same to avoid any potential bias introduced by sample size, as it is established that data-adaptive mechanisms show improved performance with larger sample sizes (Luque-Fernandez et al., 2018).

4.1 Evaluation criteria

The performance measures used to evaluate the simulation study, with the ATE ($\psi_{ATE} \equiv \psi$) as the target parameter, are as follows:

- **Bias:**
  - Definition: $\text{E}[\hat{\psi}] - \psi$
  - Estimate: $\frac{1}{n_{\text{sim}}}\sum_{i=1}^{n_{\text{sim}}}\hat{\psi}_i - \psi$
- **Relative Bias in %:**
  - Definition: $\left(\frac{\text{Bias}}{\psi}\right)100$
  - Estimate: $\left(\frac{\widehat{\text{Bias}}}{\hat{\psi}}\right)100$
- **Empirical standard error (EmpSE):**
  - Definition: $\sqrt{\text{Var}(\hat{\psi})}$
  - Estimate: $\sqrt{\frac{1}{n_{\text{sim}}-1}\sum_{i=1}^{n_{\text{sim}}}\left(\hat{\psi}_i - \bar{\psi}\right)^2}$

- **Root mean square error (RMSE):**
  - Definition: $\sqrt{\mathrm{E}\left[(\hat{\psi}-\psi)^2\right]}$
  - Estimate: $\sqrt{\frac{1}{n_{\mathrm{sim}}}\sum_{i=1}^{n_{\mathrm{sim}}}\left(\hat{\psi}_i-\psi\right)^2}$
- **Model standard error (ModSE):**
  - Definition: $\sqrt{\mathrm{E}[\widehat{\mathrm{Var}}(\hat{\psi})]}$
  - Estimate: $\sqrt{\frac{1}{n_{\mathrm{sim}}}\sum_{i=1}^{n_{\mathrm{sim}}}\widehat{\mathrm{Var}\left(\hat{\psi}_i\right)}}$
- **Error in ModSE in %:**
  - Definition: $100\left(\frac{\mathrm{ModSE}}{\mathrm{EmpSE}}-1\right)$
  - Estimate: $100\left(\frac{\widehat{\mathrm{ModSE}}}{\widehat{\mathrm{EmpSE}}}-1\right)$
- **Coverage:**
  - Definition: $P\left(\hat{\psi}_{\mathrm{low}}\ \leq\psi\leq\hat{\psi}_{\mathrm{upp}}\right)$
  - Estimate: $\frac{1}{n_{\mathrm{sim}}}\sum_{i=1}^{n_{\mathrm{sim}}}1\left(\hat{\psi}_{\mathrm{low},i}\leq\psi\leq\hat{\psi}_{\mathrm{upp},i}\right)$
- **Bias eliminated coverage:**
  - Definition: $P\left(\hat{\psi}_{\mathrm{low}}\ \leq\bar{\psi}\leq\hat{\psi}_{\mathrm{upp}}\right)$
  - Estimate: $\frac{1}{n_{\mathrm{sim}}}\sum_{i=1}^{n_{\mathrm{sim}}}1\left(\hat{\psi}_{\mathrm{low},i}\leq\bar{\psi}\leq\hat{\psi}_{\mathrm{upp},i}\right)$
- **Mean confidence interval (CI) length:**
  - Definition: $E\left(\hat{\psi}_{\mathrm{upp}}\ -\hat{\psi}_{\mathrm{low}}\right)$
  - Estimate: $\frac{1}{n_{\mathrm{sim}}}\sum_{i=1}^{n_{\mathrm{sim}}}\left(\hat{\psi}_{\mathrm{upp},i}-\hat{\psi}_{\mathrm{low},i}\right)$
- **Proportional CI length:**
  - Definition: the mean CI length of each method is considered and related to the greatest mean CI length of all methods

The mathematical formulations for these computations (despite mean CI length and proportional CI length) are detailed in the work of Morris et al. (2019).

4.2 Data generating process

In the following, the data generating processes (DGP) will be described. Firstly, the creation of data as conducted in the first version of Dashti et al. (2021). Secondly, furhter modifications like increasing positivity violation, adding continous and/or categorical confounders and incorporating some dependencies between the confounders were made to test the proposed methods in various data settings.

Reproduced data:

In Dashti et al. (2021), an attempt was made to replicate the Victorian Adolescent Health Cohort Study (VAHCS) data, which are not publicly available yet. Minor but reasonable changes were made. The main takeaways are, that parametric regression models were used to generate the data and taking into account three scenarios (simple, complex 1, and complex 2) which progressively increase in complexity due to the presence of confounder-confounder interaction terms. Particularly, these scenarios vary in terms of the exposure and outcome generation models. The simple scenario does not involve any confounder-confounder interaction terms. In contrast, the complex 1 scenario includes such interaction terms, and the complex 2 scenario contains the same interaction terms but with coefficients approximately two times larger than those in the complex 1 scenario. For all scenarios, an auxiliary variable B with a standard normal distribution was generated, along with a set of binary confounders $W = (W_1, W_2, W_3, W_4, W_5)$. Here, confounders W_2, W_3, and W_4 were generated through regression on B. The NPSEM in this case is:

$$\begin{aligned} B &= f_B\left(U_B\right), \\ W &= f_W\left(B, U_W\right), \\ A &= f_A\left(B, W, U_A\right), \\ Y &= f_Y\left(W, A, U_Y\right), \end{aligned}$$

and in specific the models for generating the variables for B and W are:

$$\begin{aligned} \mathrm{B} &\sim \mathcal{N}(0,1) \\ \mathrm{W1} &\sim \text{Binomial}\left(1, \text{logit}^{-1}\left(\alpha_0\right)\right) \\ \mathrm{W2} &\sim \text{Binomial}\left(1, \text{logit}^{-1}\left(\beta_0 + \beta_1 \mathrm{B}\right)\right) \\ \mathrm{W3} &\sim \text{Binomial}\left(1, \text{logit}^{-1}\left(\gamma_0 + \gamma_1 \mathrm{B}\right)\right) \\ \mathrm{W4} &\sim \text{Binomial}\left(1, \text{logit}^{-1}\left(\delta_0 + \delta_1 \mathrm{B}\right)\right) \\ \mathrm{W5} &\sim \text{Binomial}\left(1, \text{logit}\, -1\left(\zeta_0\right)\right). \end{aligned}$$

Furthermore it is assumed that all binary variables are coded 0/1 and $\text{logit}^{-1}((\cdot)) = \exp(\cdot)/(1+\exp(\cdot))$. In the simple scenario, a binary exposure A was generated through a main-effects logistic regression on B and W, while a continuous outcome Y was generated through a main-effects linear regression on

A and W:

$$\begin{aligned}
\mathrm{A}_{\text{simple}} &\sim \text{Binomial}\left(1, \text{logit}^{-1}\left(\eta_0 + \eta_1 \mathrm{W}_1 + \eta_2 \mathrm{W}_2 + \eta_3 W_3 + \eta_4 \mathrm{W}_4 + \eta_5 \mathrm{W}_5 + \eta_6\ \mathrm{B}\right)\right) \\
\mathrm{Y}_{\text{simple}} &\sim \mathcal{N}\left(\theta_0 + \theta_1 \mathrm{A} + \theta_2 \mathrm{W}_1 + \theta_3 \mathrm{W}_2 + \theta_4 \mathrm{W}_3 + \theta_5 \mathrm{W}_4 + \theta_6 \mathrm{W}_5, \text{sd} = 1\right)
\end{aligned}$$

In the complex scenarios A was generated through regression on B, W, and involved two-way confounder-confounder interactions. Similarly, Y was generated through regression on A, W, and encompassed two-, three-, and four-way confounder-confounder interactions. Interactions with W_2 were not included in the exposure and outcome models.

$$\begin{aligned}
\mathrm{A}_{\text{complex}} &\sim \text{Binomial}\left(1, \text{logit}^{-1}\left(\eta_0 + \eta_1 \mathrm{W}_1 + \eta_2 \mathrm{W}_2 + \eta_3 \mathrm{W}_3 + \eta_4 \mathrm{W}_4 + \eta_5 \mathrm{W}_5 + \eta_6 \mathrm{B} + \eta_7 \mathrm{W}_1 \mathrm{W}_3\right.\right. \\
&\quad \left.\left. + \eta_8 \mathrm{W}_1 \mathrm{W}_4 + \eta_9 \mathrm{W}_1 \mathrm{W}_5 + \eta_{10} \mathrm{W}_3 \mathrm{W}_4 + \eta_{11} \mathrm{W}_3 \mathrm{W}_5 + \eta_{12} \mathrm{W}_4 \mathrm{W}_5\right)\right) \\
\mathrm{Y}_{\text{complex}} &\sim \mathcal{N}\left(\theta_0 + \theta_1 \mathrm{A} + \theta_2 \mathrm{W}_1 + \theta_3 \mathrm{W}_2 + \theta_4 \mathrm{W}_3 + \theta_5 \mathrm{W}_4 + \theta_6 \mathrm{W}_5 + \theta_7 \mathrm{W}_1 \mathrm{W}_3 + \theta_8 \mathrm{W}_1 \mathrm{W}_4 + \theta_9 \mathrm{W}_1 \mathrm{W}_5\right. \\
&\quad + \theta_{10} \mathrm{W}_3 \mathrm{W}_4 + \theta_{11} \mathrm{W}_3 \mathrm{W}_5 + \theta_{12} \mathrm{W}_4 \mathrm{W}_5 + \theta_{13} \mathrm{W}_1 \mathrm{W}_3 \mathrm{W}_4 + \theta_{14} \mathrm{W}_1 \mathrm{W}_3 \mathrm{W}_5 \\
&\quad \left. + \theta_{15} \mathrm{W}_1 \mathrm{W}_4 \mathrm{W}_5 + \theta_{16} \mathrm{W}_3 \mathrm{W}_4 \mathrm{W}_5 + \theta_{17} \mathrm{W}_1 \mathrm{W}_3 \mathrm{W}_4 \mathrm{W}_5, \text{sd} = 1\right)
\end{aligned}$$

The coefficients for the respective models are depicted in table 2. Compared to Dashti et al. (2021), there was only one change concerning the intercept value for A in the complex scenario 2, which was -2.25 instead of -2.40. The range of the coefficient values for the interaction terms in the exposure and outcome models was from -1.6 to 0.3 and from -1.2 to 1.7 respectively for complex scenario 1. For complex scenario 2, these values ranged from -3.2 to 0.5 and from -2.4 to 3.4 respectively. Adjustments were made to the intercept to ensure an approximately 15% in all scenarios within the simulated data. Across all outcome generation models, the coefficient for A (θ_1), representing the true value of the ATE, was set to 0.2 . This value denotes a moderate effect size, with the null hypothesis of no causal effect being formally rejected ($\mathrm{p} < 0.05$) in around 80% of the simulated datasets. The outcome model's intercept was also modified to maintain the mean of Y at 0. The distributions of the variables and the simulated frequencies can be found in table 1. In the simple scenario, the positivity violation was $\sim$ 0.02%, in the complex scenario 1 it was $\sim$ 1.10% and in the complex scenario 2 it was $\sim$ 9.25%. An illustrative example of the probability density function of the propensity score by treatment status can be found in figure 3.

Table 1: Parameter 1 represents proportion for binary/categorical variables and mean for normal distributed outcome. Parameter 2 represents the standard deviation.

Distribution of variables in simulated complete data

		W1	W2	W3	W4	W5	A	Y
	Type	binary	binary	binary	binary	binary	binary	normal
For all scenarios	Parameter 1	0.21	0.14	0.59	0.37	0.38	0.15	0
	Parameter 2							1.1

		% with missing value						
		W2	W3	W4	A	Y	A/Y	Any
	DAG T	30	15	20	30	20	40	50
	DAG A	30	15	20	30	20	40	50
	DAG B	30	15	20	30	20	40	50
	DAG C	30	15	20	30	20	40	50
	DAG D	30	15	20	30	20	40	50
For all scenarios	DAG E	30	15	20	30	20	40	50
	DAG F	30	15	20	30	20	40	50
	DAG G	30	15	20	30	20	40	50
	DAG H	30	15	20	30	20	40	50
	DAG I	30	15	20	30	20	40	50
	DAG J	30	15	20	30	20	40	50

Distribution of variables in the simulated complete data and proportion with missingness across the 11 simulated missingness mechanisms and the different considered scenarios.

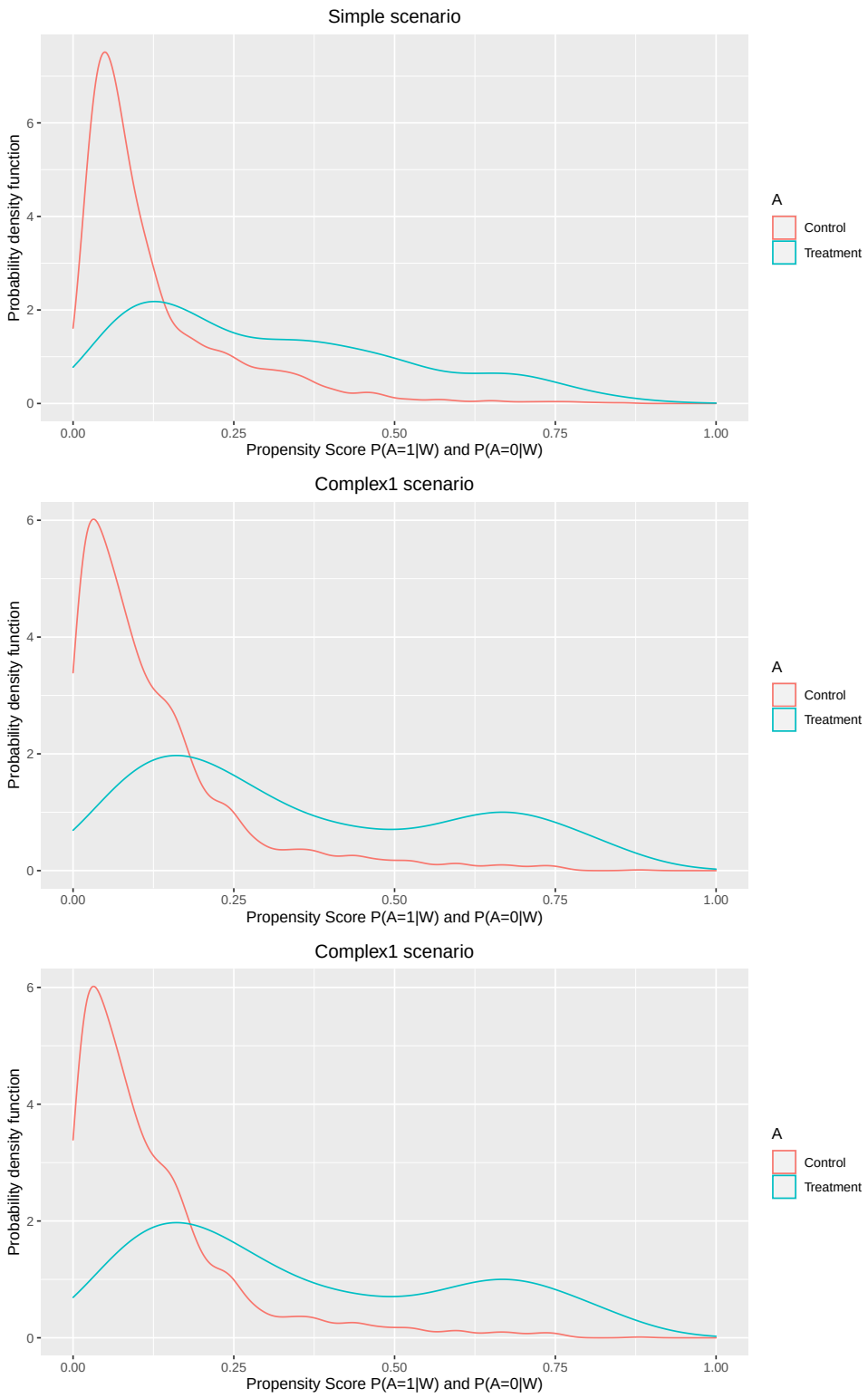


Figure 3: Probability density function of the propensity score by treatment status for one randomly selected data set from 1000 simulated data sets across the different scenarios.

	Model for	Regression coefficient of											
		Intercept	W1	W2	W3	W4	W5	A	Y	B	MW2	MW3	MW4	MZA
Complete Data	W1	-1.30												
	W2	-1.90								0.40				
	W3	0.40								0.70				
	W4	-0.60								-0.70				
	W5	-0.50												
		Simple scenario												
	A	-2.90	1.30	1.90	0.40	0.20	-0.30			0.70				
	Y	-070	0.10	0.40	0.70	0.20	0.30	0.20						
		Complex scenarios												
	A		1.30	-1.90	0.40	0.20	-0.30			0.70				
	Y		0.10	0.40	0.70	0.20	0.30	0.20						
DAG T	MW2	-0.85												
	MW3	-4.40									4.30			
	MW4	-4.00									3.90	1.40		
	MA	-2.00									1.50	1.50	1.50	
	MY	-1.60									-0.50	0.50	0.50	0.50
DAG A	MW2	-1.45	0.90				0.90							
	MW3	-5.60	0.90				0.90				4.80			
	MW4	-4.70	0.90				0.90				3.90	1.50		
	MA	-2.50	0.90				0.90				1.30	1.50	1.50	
	MY	-2.10	0.90				0.90				0.10	0.10	0.10	0.05
DAG B	MW2	-1.60	0.90				0.90	0.90						
	MW3	-5.20	0.90				0.90	0.90			4.10			
	MW4	-4.40	0.90				0.90	0.90			3.20	2.00		
	MA	-3.70	0.90	0.90	0.90	0.90	0.90				1.80	1.50	1.50	
	MY	-3.25	0.90	0.90	0.90	0.90	0.90	0.90			-0.30	0.10	0.10	0.10
DAG C	MW2	-1.60	0.90				0.90	0.90	0.10					
	MW3	-5.30	0.90				0.90	0.90	0.10		4.30			
	MW4	-4.50	0.90				0.90	0.90	0.10		3.50	1.30		
	MA	-3.70	0.90	0.90	0.90	0.90	0.90		0.10		1.70	1.50	1.50	
	MY	-3.25	0.90	0.90	0.90	0.90	0.90	0.90			-0.40	0.10	0.10	0.15
DAG D	MW2	-1.60	0.90	0.90			0.90							
	MW3	-6.20	0.90		0.90		0.90				4.80			
	MW4	-5.00	0.90			0.90	0.90				3.90	1.50		
	MA	-2.65	0.90				0.90	0.90			1.30	1.50	1.50	
	MY	-2.10	0.90				0.90				0.10	0.10	0.10	0.10
DAG E	MW2	-1.75	0.90	0.90			0.90	0.90						
	MW3	-5.70	0.90		0.90		0.90	0.90			4.10			
	MW4	-4.80	0.90			0.90	0.90	0.90			3.20	2.00		
	MA	-3.80	0.90	0.90	0.90	0.90	0.90	0.90			1.50	1.50	1.50	
	MY	-3.20	0.90	0.90	0.90	0.90	0.90	0.90			-0.60	0.10	0.10	0.20
DAG F	MW2	-1.60	0.90	0.90			0.90		0.10					
	MW3	-6.60	0.90		0.90		0.90		0.10		5.20			
	MW4	-5.40	0.90			0.90	0.90		0.10		4.20	1.70		
	MA	-2.55	0.90				0.90	0.90	0.10		1.20	1.30	1.30	
	MY	-2.10	0.90				0.90				-0.30	0.10	0.10	0.40
DAG G	MW2	-1.60	0.90				0.90	0.90						
	MW3	-5.20	0.90				0.90	0.90			4.10			
	MW4	-4.45	0.90				0.90	0.90			3.20	2.00		
	MA	-3.70	0.90	0.90	0.90	0.90	0.90				1.70	1.50	1.50	
	MY	-3.30	0.90	0.90	0.90	0.90	0.90	0.90	0.10		-0.30	0.10	0.10	0.10
DAG H	MW2	-1.60	0.90				0.90	0.90	0.10					
	MW3	-5.40	0.90				0.90	0.90	0.10		4.30			
	MW4	-4.50	0.90				0.90	0.90	0.10		3.50	1.30		
	MA	-3.65	0.90	0.90	0.90	0.90	0.90		0.10		1.50	1.50	1.50	
	MY	-3.30	0.90	0.90	0.90	0.90	0.90	0.90	0.10		-0.50	0.30	0.30	0.10
DAG I	MW2	-1.75	0.90	0.90			0.90	0.90	0.10					
	MW3	-5.95	0.90		0.90		0.90	0.90	0.10		4.30			
	MW4	-4.80	0.90			0.90	0.90	0.90	0.10		3.50	1.30		
	MA	-3.80	0.90	0.90	0.90	0.90	0.90	0.90	0.10		1.50	1.50	1.50	
	MY	-3.20	0.90	0.90	0.90	0.90	0.90	0.90			-0.60	0.10	0.10	0.20
DAG J	MW2	-1.70	0.90	0.90			0.90	0.90	0.10					
	MW3	-5.95	0.90		0.90		0.90	0.90	0.10		4.30			
	MW4	-4.85	0.90			0.90	0.90	0.90	0.10		3.50	1.30		
	MA	-3.80	0.90	0.90	0.90	0.90	0.90	0.90	0.10		1.50	1.50	1.50	
	MY	-3.35	0.90	0.90	0.90	0.90	0.90	0.90	0.10		0.05	0.05	0.05	0.05
*For the complex scenarios models for A and Y also included interactions as follows														
			W1W3	W1W4	W1W5	W3W4	W3W5	W4W5	W1W3W4	W1W3W5	W1W4W5	W3W4W5	W1W3W4W5	
		Complex 1 scenario												
	A	-2.40	-1.60	-1.20	-0.50	-0.60	0.30	-1.50						
	Y	-0.70	-0.50	1.00	0.10	0.10	0.40	-0.10	-1.20	-1.00	-0.10	-0.40	1.70	
		Complex 2 scenario												
	A	-2.25	-3.20	-2.30	-1.00	-1.20	0.50	-2.90						
	Y	-0.70	-0.90	2.00	0.20	0.20	0.70	-0.20	-2.40	-2.00	-0.30	-0.80	3.40	

Table 2: Coefficient values are displayed for data generation and also for missing indicators across all considered scenarios for reproduced Data.

Positivity violation data:

Positivity violations can be particularly problematic in observational studies, where the assignment of treatment or exposure is not under the control of the researcher. In such cases, there might be certain groups of individuals who, due to their characteristics, are never assigned one of the treatment levels (Petersen et al., 2012). Although (TMLE) is a powerful method for estimating causal effects, it's not immune to the issue of positivity violation. If certain combinations of covariates and treatment levels are not observed in the data, then the treatment model used by TMLE would incorrectly estimate a zero probability for these combinations and will lead to biased estimates of the causal effect. Moreover, some treatment assignment probabilities might be exceedingly small. As a result, the values of $H_n^*(A, W)$ could be extremely large for a subset of observations. This lack of identifiability may result in extreme weights during the estimation procedure, introducing unnecessary variance and potentially leading to instability in the estimates. To mitigate positivity violations, the application of techniques such as truncation (i.e. Bounding $g_n(A \mid W)$ away from $(0, 1)$) is recommended. However, truncation introduces bias, necessitating a trade-off between variance and bias. These effects are mitigated by fluctuating on the logit scale (the default in the used R-packages; also introduced in Chapter 3.3.2). Furthermore it is important to note, that the logistic fluctuation for continuous Y requires Y to be bounded by (a, b). When these upper and lower bounds on Y are not provided by the user, the default is to use the range of the observed outcomes. This may be problematic when there is missingness in the outcome (i.e. MNAR case) if the distribution of observed outcomes is truncated with respect to the true distribution of the outcome (Gruber and van der Laan, 2012).

For these reasons, it would be intriguing to further examine the effect of increasing the positivity violation of the reproduced data through a slight modification in the DGP. The procedure was as follows: First, the auxiliary variable B and the binary confounders $W = (W_1, W_2, W_3, W_4, W_5)$ were created as outlined in the 'Reproduced data' section. Following this, a grid-search was conducted for all 13 parameters (intercept, main effects, and interaction terms) for the exposure model, which was based on logistic regression. This process skipped the simple scenario from the 'reproduced data' section. All parameters were randomly drawn together, each from a uniform distribution between -3.5 and 3.5. For each draw, these parameters were utilized in the exposure model, and then the proportion of observations where the predicted probability of treatment was below 0.002 was calculated. This threshold is significant as values below it could lead to positivity violations in a causal inference context. The parameters that resulted in the data being most closely aligned with the conditions where 30% and 40% of observations had a predicted treatment probability below 0.002 were retained.In essence, a simplified form of a grid search was applied across randomly generated parameters. This was done to identify the parameters of a logistic regression model that could generate data with certain characteristics, specifically 30% and 40% of observations demonstrating positivity violations. For both positivity violation scenarios, the outcome variable Y was generated using regression on A, W and incorporated two-way, three-way, and four-way interactions among the confounders, as described in the 'reproduced data' section. In addition, there was a variation in complexity. For example, in the complex 2 scenario with a 30% positivity violation, the coefficients were approximately twice as large as those in the complex 1 scenario. Therefore, in total, two distinct datasets were created for each

desired level of positivity violation. The distributions of the variables and the simulated frequencies can be obtained from table 1 whereas the coefficients for the respective models are depicted in table B.1.

Modified data:

In order to evaluate the missing data method MI with **Amelia** in conjunction with TMLE, it was necessary to include also continuous confounders in the data setup. Therefore a further modification to the data in the 'Reproduced data' section was carried out by replacing the binary confounders W_4, W_5 with continuous variables, drawn from a normal distribution. The mean of W_4 was generated via regression on B. Their respective standard deviations were set to 1 and 2. These models can be represented as follows:

$$\begin{aligned}
\mathrm{W4} &\sim \mathcal{N}\left(\delta 0 + \delta_1 \mathrm{B}, \mathrm{sd} = 1\right) \\
\mathrm{W5} &\sim \mathcal{N}\left(\zeta_0, \mathrm{sd} = 2\right).
\end{aligned}$$

The models for the outcome Y and the exposure A were identical in the simple scenario. Nevertheless, in the complex scenarios, higher interactions with W_2 were incorporated. These were not included in the two-way interactions but they were a part of the three-way and four-way interactions, essentially replacing the confounder W_3. As a consequence, the model for Y is given through:

$$\begin{aligned}
\mathrm{Y}_{\mathrm{complex}} \sim \mathcal{N}\,(&\theta_0 + \theta_1 \mathrm{A} + \theta_2 \mathrm{W}_1 + \theta_3 \mathrm{W}_2 + \theta_4 \mathrm{W}_3 + \theta_5 \mathrm{W}_4 + \theta_6 \mathrm{W}_5 + \theta_7 \mathrm{W}_1\mathrm{W}_3 + \theta_8 \mathrm{W}_1\mathrm{W}_4 + \theta_9 \mathrm{W}_1\mathrm{W}_5 \\
&+ \theta_{10} \mathrm{W}_3\mathrm{W}_4 + \theta_{11} \mathrm{W}_3\mathrm{W}_5 + \theta_{12} \mathrm{W}_4\mathrm{W}_5 + \theta_{13} \mathrm{W}_1\mathrm{W}_2\mathrm{W}_4 + \theta_{14} \mathrm{W}_1\mathrm{W}_2\mathrm{W}_5 \\
&+ \theta_{15} \mathrm{W}_1\mathrm{W}_4\mathrm{W}_5 + \theta_{16} \mathrm{W}_2\mathrm{W}_4\mathrm{W}_5 + \theta_{17} \mathrm{W}_1\mathrm{W}_2\mathrm{W}_4\mathrm{W}_5, \mathrm{sd} = 1)
\end{aligned}$$

In the complex 2 scenario, the coefficients of the interaction terms in the exposure and outcome models were approximately twice as large as those in the complex 1 scenario. Additionally, unlike the approach outlined in the 'positivity violation data' section, the positivity violation was manually adjusted. In the simple scenario, the positivity violation was $\sim 0\%$, while in complex 1 scenario it was $\sim 10\%$, and in complex 1 scenario, it was $\sim 32.25\%$. Furthermore, the proportions of binary confounders W_1, W_2 and W_3 were adjusted. The distributions of the variables and the simulated frequencies can be found in table B.8 whereas the coefficients for the respective models are depicted in table B.2.

Copula 1 data:

This data setup follows the exact same procedure as described in the 'modified data' section. The only difference is the utilization of the **copula** r-package (Hofert et al., 2023) to model interdependencies between the different confounders $W = (W_1, W_2, W_3, W_4, W_5)$. In essence, a copula is a multivariate distribution function where the marginal distributions are uniformly distributed on the interval $[0, 1]$. The role of a copula is to describe the correlation or dependence structure between these variables, independently of the margins. This separation of marginals and dependence is what gives copulas their flexibility and makes them a powerful tool in multivariate modeling (Nelsen, 2006). In this DGP, the

Gaussian copula is employed to generate dependent covariates, with the following correlation matrix:

$$\rho = \begin{bmatrix} 1 & 0.3 & -0.3 & 0.3 & 0.3 \\ 0.3 & 1 & 0.7 & 0.3 & 0.3 \\ -0.3 & 0.7 & 1 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 & 1 & 0.7 \\ 0.3 & 0.3 & 0.3 & 0.7 & 1 \end{bmatrix}$$

From this, the marginal distributions are generated, including three Bernoulli and two normal marginals. The Bernoulli marginals W_1, W_2 and W_3 are created using an ifelse statement along with the corresponding logit^{-1} function from the 'reproduced data' section to define the probability of success. Thus resulting into a 1 if the sample drawn from the Gaussian copula is larger than the respective modeled probability of success and 0 otherwise. The normal marginals W_4, W_5 are generated through the normal quantile function by modeling the mean and standard deviation as stated in the 'modified data' section, where the respective quantiles for each observation are accessed through the random draws of the Gaussian copula. In the simple scenario, the positivity violation was $\sim 0.25\%$, while in complex scenario 1 it was $\sim 9.75\%$, and in complex scenario 2, it was $\sim 31.20\%$. The distributions of the variables and the simulated frequencies can be obtained in table B.3 and the coefficients for the respective models are depicted in table B.3.

Copula 2 data:

In the next DGP the binary confounder W_3 is replaced by a categorical variable with four categories $W_{3_1}, W_{3_2}, W_{3_3}, W_{3_4}$. The probabilities of each category depend on B and are created using a softmax function:

$$P(W_3 = i) = \frac{e^{\gamma_{0_i} + \gamma_{1_i} \cdot B}}{\sum_{j=1}^{4} e^{\gamma_{0_j} + \gamma_{1_j} \cdot B}}, \quad \text{for } i, j = 1, ..., 4.$$

The probabilities created in the previous step are cumulative. This means that they represent the probability of W_3 falling into category i or any category below it. These cumulative probabilities are calculated for each observation. W_3 is assigned to the respective category such that the drawn sample from the Gaussian copula falls in the range of the cumulative probability of that category. The other confounders are generated as in the 'copula 1 data' section. For the simple scenario the models for exposure and outcome are:

$$A_{\text{simple}} \sim \text{Binomial}(1, \text{logit}^{-1}(\eta_0 + \eta_1 W_1 + \eta_2 W_2 + \eta_3 W_{3_1} + \eta_4 W_{3_2} + \eta_5 W_{3_3} + \eta_6 W_4 + \eta_7 W_5 + \eta_8 B))$$
$$Y_{\text{simple}} \sim \mathcal{N}(\theta_0 + \theta_1 A + \theta_2 W_1 + \theta_3 W_2 + \theta_4 W_{3_1} + \theta_5 W_{3_2} + \theta_6 W_{3_3} + \theta_7 W_4 + \theta_8 W_5, \text{sd} = 1).$$

For the complex scenarios the models are:

$$
\begin{aligned}
A_{\text{complex}} \sim \text{Binomial}(1, \text{logit}^{-1}(&\eta_0 + \eta_1 W_1 + \eta_2 W_2 + \eta_3 W_{3_1} + \eta_4 W_{3_2} + \eta_5 W_{3_3} + \eta_6 W_4 + \eta_7 W_5 \\
&+ \eta_8 B + \eta_9 W_1 W_{3_1} + \eta_{10} W_1 W_{3_2} + \eta_{11} W_1 W_{3_3} + \eta_{12} W_1 W_4 + \eta_{13} W_1 W_5 \\
&+ \eta_{14} W_{3_1} W_4 + \eta_{15} W_{3_2} W_4 + \eta_{16} W_{3_3} W_4 + \eta_{17} W_{3_1} W_5 + \eta_{18} W_{3_2} W_5 \\
&+ \eta_{19} W_{3_3} W_5 + \eta_{20} W_4 W_5)) \\
Y_{\text{complex}} \sim \mathcal{N}(&\theta_0 + \theta_1 A + \theta_2 W_1 + \theta_3 W_2 + \theta_4 W_{3_1} + \theta_5 W_{3_2} + \theta_6 W_{3_3} + \theta_7 W_4 + \theta_8 W_5 + \theta_9 W_1 W_{3_1} \\
&+ \theta_{10} W_1 W_{3_2} + \theta_{11} W_1 W_{3_3} + \theta_{12} W_1 W_4 + \theta_{13} W_1 W_5 + \theta_{14} W_{3_1} W_4 + \theta_{15} W_{3_2} W_4 \\
&+ \theta_{16} W_{3_3} W_4 + \theta_{17} W_{3_1} W_5 + \theta_{18} W_{3_2} W_5 + \theta_{19} W_{3_3} W_5 + \theta_{20} W_4 W_5 + \theta_{21} W_1 W_2 W_4 \\
&+ \theta_{22} W_1 W_2 W_5 + \theta_{23} W_1 W_4 W_5 + \theta_{24} W_2 W_4 W_5 + \theta_{25} W_1 W_2 W_4 W_5, \text{sd} = 1).
\end{aligned}
$$

In the simple scenario, the positivity violation was $\sim 0.2\%$, while in complex scenario 1 it was $\sim 9\%$, and in complex scenario 2, it was $\sim 29.90\%$. The distributions of the variables and the simulated frequencies can be found in table B.9 and the coefficients for the respective models are depicted in table B.4.

Copula 3 data: The last DGP incorporates an additional continuous variable to increase complexity. Therefore, the correlation matrix for the Gaussian copula requires enlargement by one dimension:

$$
\rho = \begin{bmatrix}
1 & 0.3 & -0.3 & 0.3 & 0.3 & -0.3 \\
0.3 & 1 & 0.7 & 0.3 & 0.3 & 0.3 \\
-0.3 & 0.7 & 1 & 0.3 & 0.3 & 0.3 \\
0.3 & 0.3 & 0.3 & 1 & 0.7 & 0.3 \\
0.3 & 0.3 & 0.3 & 0.7 & 1 & -0.3 \\
-0.3 & 0.3 & 0.3 & 0.3 & -0.3 & 1
\end{bmatrix}
$$

Therefore a gamma-distributed variable was selected. The shape and rate parameters of the gamma distribution are linear functions of t_B, the truncated version of B with the range of $[-0.99, 0.99]$. This requirement ensures that the parameters of the gamma distribution remain positive. The equations for the parameters are shape $= \xi_0 + \xi_1 t_B$ and rate $= \tau_0 + \tau_1 t_B$. The gamma marginal of W_6 are generated through the gamma quantile function by using the modeled shape and rate parameters, where the respective quantiles for each observation are accessed through the random draws of the Gaussian copula. For the simple scenario the models for exposure and outcome are:

$$
\begin{aligned}
A_{\text{simple}} \sim \text{Binomial}(1, \text{logit}^{-1}(&\eta_0 + \eta_1 W_1 + \eta_2 W_2 + \eta_3 W_{3_1} + \eta_4 W_{3_2} + \eta_5 W_{3_3} + \eta_6 W_4 \\
&+ \eta_7 W_5 + \eta_8 W_6 + \eta_9 B)) \\
Y_{\text{simple}} \sim \mathcal{N}(&\theta_0 + \theta_1 A + \theta_2 W_1 + \theta_3 W_2 + \theta_4 W_{3_1} + \theta_5 W_{3_2} + \theta_6 W_{3_3} + \theta_7 W_4 + \theta_8 W_5 + \theta_9 W_6, \text{sd} = 1).
\end{aligned}
$$

For the complex scenarios the models are:

$$
\begin{aligned}
A_{\text{complex}} \sim \text{Binomial}(1, \text{logit}^{-1}(&\eta_0 + \eta_1 W_1 + \eta_2 W_2 + \eta_3 W_{3_1} + \eta_4 W_{3_2} + \eta_5 W_{3_3} + \eta_6 W_4 + \eta_7 W_5 \\
&+ \eta_8 W_6 + \eta_9 B + \eta_{10} W_1 W_{3_1} + \eta_{11} W_1 W_{3_2} + \eta_{12} W_1 W_{3_3} + \eta_{13} W_1 W_4 \\
&+ \eta_{14} W_1 W_5 + \eta_{15} W_{3_1} W_4 + \eta_{16} W_{3_2} W_4 + \eta_{17} W_{3_3} W_4 + \eta_{18} W_{3_1} W_5 + \eta_{19} W_{3_2} W_5 \\
&+ \eta_{20} W_{3_3} W_5 + \eta_{21} W_4 W_5 + \eta_{22} W_1 W_6 + \eta_{23} W_4 W_6 + \eta_{24} W_5 W_6)) \\
Y_{\text{complex}} \sim \mathcal{N}(&\theta_0 + \theta_1 A + \theta_2 W_1 + \theta_3 W_2 + \theta_4 W_{3_1} + \theta_5 W_{3_2} + \theta_6 W_{3_3} + \theta_7 W_4 + \theta_8 W_5 + \theta_9 W_6 \\
&+ \theta_{10} W_1 W_{3_1} + \theta_{11} W_1 W_{3_2} + \theta_{12} W_1 W_{3_3} + \theta_{13} W_1 W_4 + \theta_{14} W_1 W_5 + \theta_{15} W_{3_1} W_4 \\
&+ \theta_{16} W_{3_2} W_4 + \theta_{17} W_{3_3} W_4 + \theta_{18} W_{3_1} W_5 + \theta_{19} W_{3_2} W_5 + \theta_{20} W_{3_3} W_5 + \theta_{21} W_4 W_5 \\
&+ \theta_{22} W_1 W_6 + \theta_{23} W_4 W_6 + \theta_{24} W_5 W_6 + \theta_{25} W_1 W_4 W_6 + \theta_{26} W_1 W_5 W_6 + \theta_{27} W_1 W_4 W_5 \\
&+ \theta_{28} W_4 W_5 W_6 + \theta_{29} W_1 W_4 W_5 W_6, \text{sd} = 1).
\end{aligned}
$$

In the simple scenario, the positivity violation was $\sim$ 0.25%, while in complex scenario 1 it was $\sim$ 9.25%, and in complex scenario 2, it was $\sim$ 30.75%. The distributions of the variables and the simulated frequencies can be found in table B.10 and the coefficients for the respective models are depicted in table B.5.

4.3 Imposing missing data

Eleven missingness scenarios were considered, as defined by the m-DAGs presented in Figure 2. These causal diagrams vary based on the presence of arrows stemming from confounders, exposure, and outcome, leading to the missingness indicators for other variables or to their own missingness indicators. These m-DAGs represent all distinct missingness scenarios in point-exposure epidemiological studies, in terms of the implications of these conditional independencies for the identifiability of key parameters (Dashti et al., 2021). Missingness was imposed on W_2, W_3, W_4, A and Y by generating missingness indicators $M_{W_2}, M_{W_3}, M_{W_4}, M_A$ and M_Y, which were coded 1 if the variable was missing and 0 if observed. For the copula 3 data missingness was as well imposed on W_6 by missingness indicator M_{W_6}. In the simulation study, variables B, W_1, and W_5 were considered as fully observed. The models used for generating these missingness indicators in the reproduced data, positivity violation data, modified

data, copula 1 data and copula 2 data were as follows:

$$
\begin{aligned}
\mathrm{M_{W_2}} &\sim \text{Binomial}(1, \text{logit}^{-1}(\iota_0 + \iota_1 \mathrm{W}_1 + \iota_2 \mathrm{W}_5 + \iota_3 \mathrm{W}_2 + \iota_4 A + \iota_5 \mathrm{Y})) \\
\mathrm{M_{W_3}} &\sim \text{Binomial}(1, \text{logit}^{-1}(\kappa_0 + \kappa_1 \mathrm{W}_1 + \kappa_2 \mathrm{W}_5 + \kappa_3 \mathrm{W}_3 + \kappa_4 A + \kappa_5 \mathrm{Y} + \kappa_6 \mathrm{M_{W_2}})) \\
\mathrm{M_{W_4}} &\sim \text{Binomial}(1, \text{logit}^{-1}(\lambda_0 + \lambda_1 \mathrm{W}_1 + \lambda_2 \mathrm{W}_5 + \lambda_3 \mathrm{W}_4 + \lambda_4 A + \lambda_5 \mathrm{Y} + \lambda_6 \mathrm{M_{W_2}} + \lambda_7 \mathrm{M_{W_3}})) \\
\mathrm{M}_A &\sim \text{Binomial}(1, \text{logit}^{-1}(\nu_0 + \upsilon_1 \mathrm{W}_1 + \nu_2 \mathrm{W}_5 + \nu_3 \mathrm{W}_2 + \nu_4 \mathrm{W}_3 + \nu_5 \mathrm{W}_4 + \nu_6 A + \nu_7 \mathrm{Y} \\
&\quad + \nu_8 \mathrm{M_{W_2}} + \nu_9 \mathrm{M_{W_3}} + \nu_{10} \mathrm{M_{W_4}})) \\
\mathrm{M_Y} &\sim \text{Binomial}(1, \text{logit}^{-1}(\xi_0 + \xi_1 \mathrm{W}_1 + \xi_2 \mathrm{W}_5 + \xi_3 \mathrm{W}_2 + \xi_4 \mathrm{W}_3 + \xi_5 \mathrm{W}_4 + \xi_6 A + \xi_7 \mathrm{Y} + \xi_8 \mathrm{M_{W_2}} \\
&\quad + \xi_9 \mathrm{M_{W_3}} + \xi_{10} \mathrm{M_{W_4}} + \xi_{11} \mathrm{M}_A))
\end{aligned}
$$

In each case, the coefficient values for confounders without missing data W_1, W_5, confounders with missing data W_2, W_3, W_4, and the exposure A were set to 0 if there was no arrow leading from the variable to the missingness indicator. If there was an arrow, the coefficients were set to 0.9, equivalent to an odds ratio (OR) of 2.5. For the outcome (Y) a similar approach was taken. In the absence of an arrow from Y to the missingness indicator, the coefficient was set to 0. If there was an arrow, the coefficient was set to 0.1, corresponding to an OR of 1.1 for a single increment increase in Y on a standardised scale, as per Dashti et al. (2021) guidelines. As outlined in the models above, the regression model for M_{W_3} incorporated the missingness indicator M_{W_2}. Similarly, the model for M_{W_4} included M_{W_2} and M_{W_3}, the model for M_A incorporated M_{W_2}, M_{W_3}, and M_{W_4}, and the model for M_Y included all the preceding missingness indicators. Hence violating the forth assumption for m-DAGs (Moreno-Betancur et al., 2018). The models used for generating the missingness indicators undergo minor adjustments when W_6 is incorporated into copula 3 data :

$$
\begin{aligned}
\mathrm{M_{W_6}} &\sim \text{Binomial}(1, \text{logit}^{-1}(\phi_0 + \phi_1 \mathrm{W}_1 + \phi_2 \mathrm{W}_5 + \phi_3 \mathrm{W}_6 + \phi_4 A + \phi_5 \mathrm{Y} + \phi_6 \mathrm{M_{W_2}} + \phi_7 \mathrm{M_{W_3}} \\
&\quad + \phi_8 M_{W_4})) \\
\mathrm{M}_A &\sim \text{Binomial}(1, \text{logit}^{-1}(\nu_0 + \upsilon_1 \mathrm{W}_1 + \nu_2 \mathrm{W}_5 + \nu_3 \mathrm{W}_2 + \nu_4 \mathrm{W}_3 + \nu_5 \mathrm{W}_4 + \nu_6 \mathrm{W}_6 + \nu_7 A + \nu_8 \mathrm{Y} \\
&\quad + \nu_9 \mathrm{M_{W_2}} + \nu_{10} \mathrm{M_{W_3}} + \nu_{11} \mathrm{M_{W_4}} + \nu_{12} \mathrm{M_{W_6}})) \\
\mathrm{M_Y} &\sim \text{Binomial}(1, \text{logit}^{-1}(\xi_0 + \xi_1 \mathrm{W}_1 + \xi_2 \mathrm{W}_5 + \xi_3 \mathrm{W}_2 + \xi_4 \mathrm{W}_3 + \xi_5 \mathrm{W}_4 + \xi_6 \mathrm{W}_6 + \xi_7 A + \xi_8 \mathrm{Y} \\
&\quad + \xi_9 \mathrm{M_{W_2}} + \xi_{10} \mathrm{M_{W_3}} + \xi_{11} \mathrm{M_{W_4}} + \xi_{12} \mathrm{M_{W_6}} + \xi_{13} \mathrm{M}_A))
\end{aligned}
$$

On top of that in the modified data, copula 1 data and copula 2 data W_4 and W_5 are normal distributed, while in copula 2 data, W_6 follows a gamma distribution. As a result, all continous confounders are standardized before they are integrated in the missing models to ensure a similar effect as the outcome induces to the respective missing indicators. If there is an arrow pointing to a missing indicator, the coefficient is set to 0.1 and otherwise it is set to 0. For both copula 2 data and copula 3 data, W_3 is a categorical variable. For these data setups, the categories W_{3_1} and W_{3_3} are utilized in the

missing models. Similar to the previous case, when an arrow exists, the coefficient is set to 0.9 and 0 otherwise. The coefficient values for these missingness indicators and the intercepts were adjusted to maintain a consistent proportion of missing data across all scenarios for each variable. For the outcome the proportion with missing data was set to 20%. The proportion with missing data for any of the outcome or exposure variables was set at 40%, while the proportion with missing data for any variable used in the target analysis (including exposure, confounders, and outcome) was set at 50%. However, the proportion of missing values in the confounders deviates in the reproduced data and positivity data table 1 from the other four data setups. The missing portions for modified data and copula 1 data are displayed in Table B.8. Similarly, those for copula 2 data are presented in Table B.9, and those for copula 3 data can be found in Table B.10.

4.4 Methods for handling missing data

In the following section, the various methods for handling missing data when estimating the ATE using TMLE, are introduced. These are broadly categorised under non-MI and MI methods and were mostly adopted from Dashti et al. (2021).

4.4.1 Non-MI approaches to handle missing data

- **Complete-case analysis (CC):**
  The most straightforward method for dealing with missing data is a complete case analysis. This method omits observations with missing data for any of the variables in the analysis. The analysis then only includes records with complete data. While simple, this approach can lead to bias, depending on the mechanism behind the missingness, and a loss of precision (White and Carlin, 2010). This approach should lead to unbiased estimation for the ATE in m-DAGS T, A, B, D, E and biased for F, G, H, I, J.

- **Extended TMLE (Ext):**
  Another strategy is the extended TMLE method to address missingness in the outcome, where records with missing exposure or confounder data are excluded. Initially, the model for $\bar{Q}_0(A, W)$ is estimated among records with complete W, A, Y data. Then, the outcome predictions are updated in the targeting step using information from both the model fitted for $g_0(1 \mid W)$ and a model fitted for $\mathrm{P}(M_Y = 0 \mid A, W)$ (the probability of having an observed outcome conditional on the exposure and confounders) among records with complete W and A data. The clever covariates $H_n^*(A, W)$ are multiplied with the inverse of $\mathrm{P}(M_Y = 0 \mid A, W)$. Updated predictions for the outcome under exposure and no exposure are thus obtained for all records, regardless of their missing outcome status. These are subsequently incorporated into the g-formula to estimate the ATE. Similar to the exposure and outcome models, the model for M_Y utilizes the same Super Learner library (Gruber and van der Laan, 2012). Provided there's no incomplete exposure and confounders, the extended TMLE method has been demonstrated to be unbiased

under an extended exchangeability assumption (namely, $Y^A \perp M_Y \mid A, W$ and $Y^A \perp A \mid W$ for $A = 0, 1$) (Diaz and van der Laan, 2017). This approach should lead to unbiased estimation for the ATE in m-DAGS T, A, B, D, E and biased for F, G, H, I, J. Since extended TMLE uses data including those with missing outcomes, it may provide more precise estimates compared to CC.

- **Extended TMLE plus missing covariate missing indicator approach (Ext MCMI):** The third method, which combines the extended TMLE approach to address missing outcome data with missing covariate missing indicators (MCMI) approach to manage missing confounder data, was utilized. This involves the inclusion of missingness indicators for the incomplete confounders in the confounding adjustment set, with records having missing exposure data being omitted. For binary or categorical incomplete confounders this is equivalent to adding an extra 'missing' category to the variable, whereas for a continuous confounder, missing observations are replaced with a fixed value, here 0.An unbiased estimate of the ATE can be yielded by the MCMI approach under an extended exchangeability assumption, $\mathrm{Y}^A \perp A \mid W, M_W$ for $A = 0, 1$, with M_W being the vector of missingness indicators for the incomplete confounders. Additionally the assumptions that the exposure or outcome only depends on the confounder when the confounder is observed, named conditionally independent treatment (CIT) or conditionally independent outcomes (CIO) (Blake et al., 2020). This approach should lead to unbiased estimation for the ATE in m-DAGS T, A, B. For m-Dags D and E the assumptions CIT and CIO are violated.

4.4.2 MI approaches to handle missing data

Several MI methods exist within MICE framework to concurrently address missing exposure, confounder, and outcome data. MICE works by specifying univariate models for each incomplete variable conditional on other variables in the imputation model, sequentially drawing imputations until convergence. In this simulation study the default of five cycles in the **mice** package was used. The **Amelia** package, on the other hand, uses a JM approach and assumes the hypothetically complete data to follow a multival normal distribution. This method estimates the parameters of the multivariate normal distribution using an expectation-maximization with bootstrapping algorithm. In **Amelia** the default for drawing until convergence was used. Both procedures are repeatedly executed to generate a total of five complete datasets. In practice, typically more than just five datasets would be imputed. However, due to computational constraints, it is limited to five in this case. Afterwards, TMLE is carried out within each complete dataset and the results are consolidated to provide the final MI estimate of the ATE and its standard error (SE) using Rubin's rules as outlined in chapter 3.1.2. In MI methods, it's recommended that all variables in the target analysis (i.e., exposure, outcome, confounders), and auxiliary variables (B), should be included in the imputation model, i.e., as predictors in each univariate model (White et al., 2010).

- **Parametric MI using linear regression (MI Reg):** This method employs logistic regression to perform multiple imputation of binary exposure and

confounders. It uses polytomous logistic regression for categorical confounder, and linear regression for continuous confounders, with the exception of W_6. For W_6, predictive mean matching (PMM), a semi-parametric imputation method, is used, as the **mice** package does not offer an imputation method specifically designed for a gamma distributed variable. Linear regression is also used for the continuous outcome. However, no interaction terms are incorporated into the univariate models.

- **Parametric MI using PMM (MI PMM):**
  This approach is characterized by its specification of univariate models for the confounders and the binary exposure, much like the previous one. However, PMM is employed this time to execute the multiple imputation of the outcome, with imputed values selected from the closest observed values post the application of a linear regression. This method was favored for its ability to address nonlinear associations, comparable to techniques such as classification and regression trees (CART) or Random Forest (RF). Within the **mice** package, the default count of donors (i.e., the number of complete cases devoid of missing values used for imputation of a missing value) is set to five. A simple random draw is employed from the pool of potential donors without any additional weighting based on the precise distance between the predicted and observed values (van Buuren, 2018). One possible issue with PMM is that it doesn't favor closer matches over others, as long as the matches come from the same donor pool. (Morris et al., 2014)

- **Parametric MI with two-way interactions (MI 2Int):**
  This approach, much like the previous one, is characterized by its specification of univariate models for the confounders, binary exposure and outcome. But this time it incorporates all two-way interactions that were used in the respective data generation (see chapter 4.2). Furthermore it includes the interactions between the exposure and outcome, exposure and each confounder, and each confounder and outcome (i.e., $AY, AW_1, YW_1, AW_3, YW_3, AW_4, YW_4, AW_5, YW_5, AW_6, YW_6$). These interaction terms are imputed using the 'passive' approach from the **mice** package and are generated within each cycle of the MI algorithm based on the current values of relevant variables involved in the interaction term. However, to facilitate the inclusion of interaction terms within each cycle, it was necessary to incorporate further auxiliary variables and manually dummy coding for the categorical variable, as this process no longer seemed to be supported (van Buuren and Groothuis-Oudshoorn, 2011). In the case that W_3 is a categorical variable, the 4th category, namely W_{3_4}, is not taken into account for the interactions, mirroring the treatment in the corresponding DGPs.

- **Parametric MI with two-, three-, and four-way interactions (MI 3Int):**
  The method carries out all two-way interactions described above (Parametric MI with two-way interactions (MI 2Int)) and further extended in this approach to additionally all include three- and four-way interactions between the confounders, that are again used in each respective DGP.

- **Parametric MI with defaults using two-way interactions (MI 2IntN):**
  This approach uses the defaults of **mice**. Meaning PMM instead of logistic regression to impute

the binary exposure/confounders, since they are of type numeric and 0/1 coded. PMM to impute all continous confounders including the outcome. Polytomous logistic regression is applied for the categorical confounder. In addition all two-way interactions that were used in the method MI 2Int were added.

- **Parametric MI with defaults using two-, three-, and four-way interactions (MI 3IntN):**
  This approach uses the defaults of **mice**. Meaning PMM instead of logistic regression to impute the binary exposure/confounders. PMM to impute all continous confounders including the outcome and polytomous logistic regression is applied for the categorical confounder. In addition all two-, three- and four-way interactions that were used in the method MI 3Int were added.

- **MI using classification and regression trees (MI CART):**
  For the next method, all variables with missing data are multiply imputed using a recursive partitioning technique, CART. It has been proposed to enable imputation that can more flexibly allow for interactions and nonlinearities. This approach allows for the imputation of missing values for specific variables by fitting a decision tree. All other variables in the imputation process are used as predictors, partitioning the data into groups based on the values of these predictors. Each record is designated to a leaf node in the tree. A donor leaf is the leaf node to which a particular record is assigned. Once this leaf is determined, a randomly chosen value from the observed values within that leaf is used to impute the missing value for the given record (Doove et al., 2014). Again the default settings of the **mice** package for the hyperparameters are used, complexity parameter = 0.0001 and minimum leaf size = 5 for CART like in Dashti et al. (2021).

- **MI using random forest (MI RF):**
  In the MI method using RF, a multitude of bootstrap samples are extracted from the complete dataset, with each sample used to fit a unique tree. Every tree contributes a 'donor leaf', and a value for the variable is randomly selected from the array of these donor leaves. It enables imputation that can more flexibly allow for interactions and nonlinearities (Doove et al., 2014). The number of trees to grow = 10 for RF in the default setting in **mice**.

- **MI using Amelia (MI Amelia):**
  For this method data preprocessing is necessary. Specifically, categorical and binary variables should be converted into factor data types because **Amelia** treats all non factor variables as numerical. Furthermore, the *noms* argument was used to ensure these variables are treated as nominal categorical variables, rather than ordinal categorical ones. Normal distributed confounders and outcomes do not require preprocessing. Moreover, the variable W_6 underwent manually logarithmization to approximate a more normal distribution, and was subsequently rescaled through the application of the exponential function (Honaker et al., 2011).

4.5 R implementations and settings for TMLE

TMLE can be utilized either through available R-packages or by manually computing the algorithm in six steps, as illustrated in (Luque-Fernandez et al., 2018). The methodology of this approach is also explained in Chapter 3.3.2. For this simulation study, three different R-packages: **tmle** (Gruber and van der Laan, 2012), **tmle3** (Coyle, 2023), and **ltmle** (Lendle et al., 2017), were used and compared. In order to ensure comparability, efforts were made to initiate a similar estimation process through appropriate argument inputs, when needed. The same Super Learner library from the package **SuperLearner** (Polley et al., 2021) was used for both estimates $g_n(1 \mid W)$ and $\bar{Q}_n^0(A, W)$ across all TMLE estimation functions. The number of folds for the cross-validation (CV) used for the Super Learner was ensured to be 5, along with truncation on $g_n(1 \mid W)$. In specific symmetrical upper and lower bounds on predicted conditional treatment assignment probabilities (propensity scores) were established at $(0.01; 0.99)$.

For the **tmle** package, two different settings were used: The first estimation process, named *TMLE*, had default settings as described in the paper by Dashti et al. (2021), with the exception of the *gbound* argument, which was set to 0.01, indicating symmetrical upper and lower bounds on the propensity scores. The second, *TMLE_S2*, used the same arguments as *TMLE* but additionally with the argument *cv.Qinit* set to FALSE and carries out the methodology from chapter 3.3.2. As consequence in *TMLE* an addional layer of CV is used to make predictions $\bar{Q}_{n,v}^0(A, W)$ on the validation sets but conditional treatment assignment probabilities were predicted on the whole data set $g_n(A \mid W)$. The remaining estimation process remains the same, since the predictions $\bar{Q}_{n,v}^0(A, W)$ on the validation sets are stacked before the updating step. The *cv.Qinit* argument is implemented to prevent overfitting $\bar{Q}_n^0(A, W)$, as overfitting could reduce the signal in the residuals required for bias reduction. Hence it is designed to protect against overfitting by cross-validating the initial Super Learner estimate of $\bar{Q}_0(A, W)$ (Gruber and van der Laan, 2012).

In the context of the **ltmle** package, only two arguments were specifically configured: For the *LTMLE* function, the V parameter was set to 5, indicating that the CV folds for Super Learner were configured to 5. In the case of the *LTMLE_ic* function, besides the V argument, the *variance.method* was set to 'ic'. This setting instructs the function to compute the variance estimate based on the influence curve. Without this configuration, the function would compute both the robust variance estimate using TMLE (an approach which directly targets the variance of the influence function as a counterfactual mean outcome (Tran et al., 2018)) and the influence curve-based variance estimate, choosing the larger of the two results for use. The primary distinction in the estimation process between the **ltmle** package and the **tmle** package is that the clever covariates $H_n^*(A, W)$ are used as weights in the logistic regression for updating $\bar{Q}_n^1(A, W)$ in the **ltmle** package. Moreover the Super Learner is slightly distinct incorporated, yielding diverse selection and weights for the learning algorithms contained in the library. This is particularly applicable when Y is continuous and needs to be rescaled to the interval $[0, 1]$.

The use of the **tmle3** package is slightly more complex, as it necessitates specifying different functions in separate steps and works in collaboration with the **sl3** package (Coyle et al., 2023) . More-

over, it utilizes 'R6' classes rather than 'S3' and 'S4' as do **tmle** and **ltmle**. Two major versions were distinguished: The first one, *CVTMLE3*, uses the TMLE methodology introduced in conjunction with CV, as outlined in Chapter 3.3.3. The estimates $g_{n,v}(A \mid W)$ and $\bar{Q}^0_{n,v}(A, W)$ on the validation sets are stacked on an earlier stage (presented in step (b)). The update is expressed as $\text{logit}\left(\bar{Q}^1_n(A, W)\right) = \text{logit}\left(\bar{Q}^0_n(A, W)\right) + \epsilon_n H^*_n(A, W)$, where the v -index dropped due to concatenation. In this scenario, the epsilon estimated for the update $\bar{Q}^1_n(A, W)$ is pooled over the V folds. It is then simply required to compute $\frac{1}{n}\sum_{i=1}^{n}\left[\bar{Q}^1_n(1, W_i) - \bar{Q}^1_n(0, W_i)\right]$. The number of folds assigned for both CVTMLE and Super Learner is set to 5. The other version *TMLE3* employs the TMLE methodology as described in chapter 3.3.2. Both versions truncate the propensity score $(0.01; 0.99)$ and use the same Super Learner library from **SuperLearner** package. Then again there are noticeable differences in comparison to the **tmle** and **ltmle** packages. Firstly, in **tmle** and **ltmle**, the outcome variable is scaled to $[0, 1]$ before estimating $\bar{Q}^0_n(A, W)$. In contrast, *TMLE3* scales Y just before it is used as an offset for the updating step, not prior to estimating $\bar{Q}^0_n(A, W)$. Furthermore CV is executed by the **sl3** package for each Super Learner fit, rather than the **SuperLearner** package as used in **tmle** and **ltmle**. As for CV-TMLE, it incorporates a nested Super Learner instead of an additional layer of CV. This procedure employs the 'Split Sequential Super Learner' method proposed by Coyle (2017), which merges folds in a manner designed to enhance speed performance while maintaining innocuousness. Further **tmle** and **tmle3** can handle missingness in the outcome through the inverse probability of censoring weights. Despite this capability, the missing data methods Ext and Ext MCMI are not executed via **tmle3**, This is due to the fact that an extra specification of the likelihood is needed for these methods to match the unified properties listed above. Unfortunately, this specification is not supported when using the inverse probability of censoring weights approach.

Generally, it is recommended to first implement the Super Learner on the full dataset, using a diverse set of algorithms. These algorithms can include for instance generalized linear models, generalized additive models, regression splines, random forests and extreme gradient boosting. Despite that, it is recommended to include simpler algorithms alongside more complex, data-adaptive ones in the Super Learner ensemble. This can provide balance and robustness, helping to improve the predictive performance of the ensemble (Balzer and Westling, 2021). Such a set of learning algorithms is summarized as 'library'. An overview of the performance of the proposed libraries in complete reproduced data (as done in Dashti et al. (2021)) is provided in Table 3. The 'lib2' library was selected for the Super Learner due to its good performance (especially bias) across all scenarios and its relatively short runtime for the scenarios and was thus used for all TMLE implementations in this simulation study. Furthermore, the positive influence becomes apparent, especially in the more complex scenarios, when Super Learner is combined with TMLE compared to simple parametric estimation without interactions.

	Estimated mean	Bias	Relative Bias (%)	empSE	RMSE	ModSE	Error in ModSE (%)	Coverage	Bias eliminated coverage	Mean CI length	Proportional CI length	Runtime 1 Core
Simple scenario												
main effects	0.20	0.00	1.01	0.08	0.08	0.08	0.62	0.95	0.95	0.32	1.00	0.27
lib 1	0.20	0.00	0.92	0.08	0.08	0.08	-3.17	0.94	0.94	0.30	0.94	8.39
lib 2	0.20	0.00	1.07	0.08	0.08	0.08	-2.92	0.94	0.94	0.30	0.94	8.36
lib 3	0.21	0.01	3.16	0.11	0.11	0.08	-28.01	0.91	0.91	0.30	0.95	10.20
lib 4	0.20	0.00	1.03	0.08	0.08	0.08	-2.98	0.94	0.94	0.30	0.95	13.30
lib 5	0.20	0.00	1.02	0.08	0.08	0.08	-8.94	0.92	0.92	0.29	0.91	31.60
lib 6	0.20	0.00	1.01	0.08	0.08	0.08	-1.83	0.95	0.95	0.30	0.96	4.65
lib 7	0.20	0.00	1.12	0.08	0.08	0.08	-2.62	0.94	0.94	0.30	0.95	4.87
lib 8	0.20	0.00	1.06	0.08	0.08	0.08	-1.30	0.94	0.94	0.30	0.96	7.61
lib 9	0.20	0.00	1.01	0.08	0.08	0.08	-1.82	0.94	0.94	0.30	0.96	1.41
Complex 1 scenario												
main effects	0.27	0.07	32.95	0.08	0.11	0.08	-0.88	0.86	0.95	0.32	1.00	0.25
lib 1	0.20	0.00	1.57	0.08	0.08	0.08	-10.30	0.92	0.92	0.29	0.91	9.41
lib 2	0.20	0.00	1.05	0.08	0.08	0.07	-9.98	0.92	0.92	0.29	0.90	9.46
lib 3	0.21	0.01	7.12	0.09	0.10	0.08	-20.22	0.88	0.88	0.29	0.91	12.40
lib 4	0.20	0.00	0.88	0.08	0.08	0.08	-9.20	0.92	0.92	0.30	0.92	21.90
lib 5	0.20	0.00	1.02	0.08	0.08	0.08	-8.94	0.92	0.92	0.29	0.91	29.18
lib 6	0.21	0.01	2.85	0.09	0.09	0.08	-9.24	0.92	0.92	0.30	0.94	4.60
lib 7	0.20	0.00	1.47	0.08	0.08	0.08	-9.21	0.92	0.92	0.30	0.92	5.02
lib 8	0.21	0.01	2.89	0.09	0.09	0.08	-8.31	0.93	0.92	0.31	0.95	15.90
lib 9	0.21	0.01	2.52	0.09	0.09	0.08	-9.01	0.92	0.92	0.30	0.95	1.85
Complex 2 scenario												
main effects	0.40	0.20	100.68	0.11	0.23	0.11	-2.11	0.54	0.93	0.43	1.00	0.23
lib 1	0.22	0.02	9.54	0.10	0.10	0.08	-20.29	0.87	0.88	0.30	0.70	9.37
lib 2	0.20	0.00	1.95	0.10	0.10	0.08	-19.29	0.88	0.88	0.30	0.70	8.94
lib 3	0.22	0.02	10.74	0.11	0.11	0.08	-26.04	0.84	0.85	0.30	0.70	11.21
lib 4	0.20	0.00	2.20	0.10	0.10	0.08	-18.25	0.89	0.89	0.30	0.71	26.00
lib 5	0.20	0.00	2.23	0.10	0.10	0.08	-18.97	0.88	0.88	0.30	0.70	32.76
lib 6	0.23	0.03	17.39	0.11	0.11	0.08	-20.66	0.85	0.87	0.32	0.76	4.31
lib 7	0.20	0.00	2.24	0.10	0.10	0.08	-18.42	0.88	0.88	0.30	0.71	4.65
lib 8	0.23	0.03	14.54	0.10	0.11	0.08	-19.20	0.86	0.88	0.32	0.76	21.10
lib 9	0.24	0.04	17.86	0.11	0.11	0.08	-21.26	0.84	0.87	0.32	0.76	1.89

Table 3: Results on complete reproduced data across all sceanrios and different library settings using the implementation *TMLE* as decribed in chapter 4.5; 'main effects' uses parametric linear regression for both exposure and outcome instead of Super Learner. The libraries are:
lib1 ← c('SL.mean','SL.glm','SL.glm.interaction', 'SL.bayesglm', 'SL.gam', 'SL.glmnet', 'SL.earth','SL.rpart', 'SL.ranger')
lib2 ← c('SL.mean','SL.glm','SL.glm.interaction', 'SL.bayesglm', 'SL.gam', 'SL.glmnet', 'SL.earth','SL.rpart', 'SL.rpartPrune', 'SL.ranger')
lib3 ← c('SL.mean','SL.glm','SL.glm.interaction', 'SL.bayesglm', 'SL.gam', 'SL.glmnet', 'SL.earth','SL.rpart', 'SL.rpartPrune', 'SL.ranger','SL.nnet')
lib4 ← c('SL.mean','SL.glm','SL.glm.interaction', 'SL.bayesglm', 'SL.gam', 'SL.glmnet', 'SL.earth','SL.rpart', 'SL.rpartPrune', 'SL.ranger','SL.step', 'SL.step.interaction')
lib5 ← c('SL.mean','SL.glm','SL.glm.interaction', 'SL.bayesglm', 'SL.gam', 'SL.glmnet', 'SL.earth','SL.rpart', 'SL.rpartPrune', 'SL.ranger','SL.randomForest')
lib6 ← c('SL.mean','SL.glm','SL.glm.interaction', 'SL.bayesglm', 'SL.gam', 'SL.glmnet', 'SL.earth')
lib7 ← c('SL.mean','SL.glm','SL.glm.interaction', 'SL.bayesglm', 'SL.gam', 'SL.glmnet', 'SL.earth', 'SL.rpartPrune')
lib8 ← c('SL.glm', 'SL.step.interaction', 'SL.earth', 'SL.mean')
lib9 ← c('SL.glm', 'SL.glm.interaction', 'SL.earth', 'SL.mean')

5 Results

5.1 Results of reproduced data

In this section, the results regarding the reproduced data will be presented. For the proposed MI methods and the implementation *TMLE* (chapter 4.5), the results of the simulation study (Dashti et al., 2021) were consistently replicated.

5.1.1 Pre-analysis

	Estimated mean	Bias	Relative Bias (%)	empSE	RMSE	ModSE	Error in ModSE (%)	Coverage	Bias eliminated coverage	Mean CI length	Proportional CI length	Runtime 1 Core
Simple scenario												
TMLE	0.20	0.00	1.00	0.08	0.08	0.08	-2.76	0.94	0.94	0.30	0.98	12.86
TMLE_S2	0.20	0.00	1.07	0.08	0.08	0.08	-3.77	0.94	0.94	0.29	0.97	6.93
LTMLE	0.21	0.01	4.93	0.08	0.08	0.08	-1.65	0.94	0.95	0.30	0.99	51.18
LTMLE_ic	0.21	0.01	4.87	0.08	0.08	0.08	-1.85	0.94	0.95	0.30	0.99	50.40
CVTMLE3	0.20	0.00	1.11	0.08	0.08	0.08	-1.62	0.94	0.94	0.30	1.00	24.83
TMLE3	0.20	0.00	1.30	0.08	0.08	0.08	-4.68	0.94	0.94	0.29	0.96	21.12
Complex 1 scenario												
TMLE	0.20	0.00	0.77	0.08	0.08	0.07	-9.60	0.92	0.92	0.29	0.98	9.96
TMLE_S2	0.20	0.00	0.65	0.08	0.08	0.07	-11.34	0.91	0.91	0.28	0.95	5.61
LTMLE	0.21	0.01	3.26	0.08	0.08	0.08	-8.02	0.92	0.93	0.29	0.99	45.93
LTMLE_ic	0.21	0.01	3.39	0.08	0.08	0.08	-8.63	0.93	0.93	0.29	0.99	45.41
CVTMLE3	0.21	0.01	3.62	0.09	0.09	0.08	-10.55	0.91	0.92	0.30	1.00	24.64
TMLE35	0.21	0.01	5.28	0.08	0.08	0.07	-16.13	0.89	0.89	0.28	0.92	21.34
Complex 2 scenario												
TMLE	0.20	0.00	1.87	0.10	0.10	0.08	-18.66	0.89	0.89	0.30	0.96	9.58
TMLE_S2	0.20	0.00	0.73	0.09	0.09	0.07	-19.48	0.88	0.88	0.29	0.92	5.59
LTMLE	0.23	0.03	15.69	0.10	0.10	0.08	-17.91	0.87	0.89	0.31	1.00	46.30
LTMLE_ic	0.23	0.03	15.81	0.10	0.10	0.08	-17.74	0.88	0.88	0.31	1.00	44.00
CVTMLE3	0.23	0.03	16.83	0.10	0.11	0.08	-24.48	0.84	0.85	0.30	0.96	24.86
TMLE3	0.24	0.04	20.17	0.10	0.11	0.07	-31.84	0.78	0.81	0.27	0.86	21.49

Table 4: Results for various TMLE implementations on complete reproduced data across all scenarios

In table 4, the different TMLE implementations for the reproduced data are shown across all scenarios. In the simple scenario, the implementations *TMLE*, *TMLE_S2*, *CVTMLE3*, and *TMLE3* have a very low bias. For *LTMLE* and *LTMLE_ic*, the relative bias is slightly higher, at about 5%. For the other performance measures, the different TMLE implementations are very similar. An exception is the slightly larger error in Model SE for *TMLE3*. In the complex 1 scenario, the relative bias for *TMLE3* and *CVTMLE3* slightly increases by 4 and 2 percentage points, respectively. Additionally, the error in Model SE increases for all implementations, especially for *TMLE3*. This also directly reflects in lower coverage. Under the complex 2 scenario, the first notable differences in performance between the individual TMLE implementations arise. The estimation of ATE remains nearly unbiased for *TMLE* and *TMLE_S2*. For *LTMLE* and *LTMLE_ic*, the relative bias is approximately 16%, while it is 17% for *CVTMLE3* and 20% for *TMLE3*. The empirical standard error (empSE) slightly increases for all implementations, with *TMLE_S2* having the lowest value, along with the root mean square

error (RMSE). The error in ModSE significantly increases for all implementations, particularly for *CVTMLE3* and *TMLE3*, which is reflected in the lower coverage. However, the bias eliminated coverage shows that the bias also increasingly affects the coverage. Furthermore, looking at the mean CI length, it can be observed that *TMLE_S2*, along with *TMLE3*, has the smallest CIs. Additionally, the implementations using the **tmle** package have the shortest runtime. For one dataset using a single core, *TMLE_S2* only requires approximately 6 seconds. The **ltmle** package takes around 9 times longer for both implementations. *TMLE3* from the **tmle3** package requires 3.5 times longer, and *CVTMLE3* takes 4 times longer. In conclusion, *TMLE_S2* offers the best overall performance combined with the lowest runtime.

5.1.2 Bias

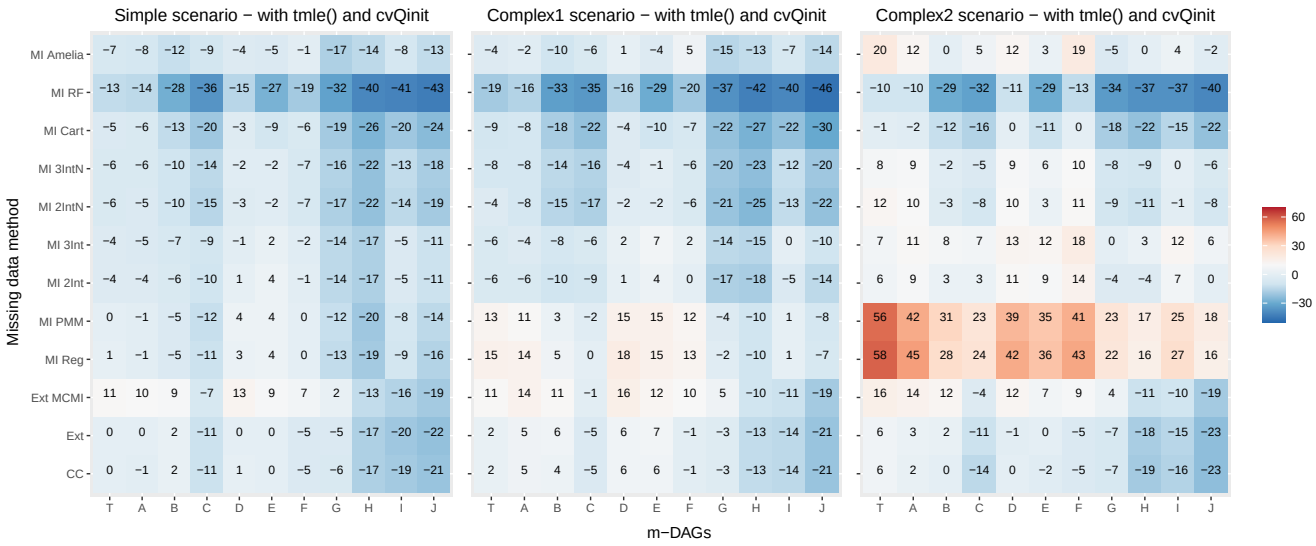


Figure 4: Relative bias (%) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE* as described in chapter 4.5 in reproduced data.

In figure 4, the relative bias of the estimated ATE of the presented missing data methods in combination with *TMLE* is shown.

For the non-MI methods (CC and Ext) under the simple scenario, the relative bias was very low for the m-DAGs T, A, B, D, E, F, G (less than 6%) for m-Dag C it was somewhat higher (-11%) and highest for m-Dags H, I, J (-17% to -22%). For the Ext MCMI method, the relative bias was significantly higher, 7-12 percentage points compared to CC and Ext methods, for the m-DAGs T,A,B,D,E, however, it was 3-4 percentage points less biased for the m-DAGs C, H, I, J. These observations are roughly reflected in the more complex scenarios (complex 1 and complex 2) as well. Here, the relative bias for the m-DAGs H, I, J remains constant, i.e., it does not tend to increase slightly as it does for the other m-DAGs.

The methods MI Reg and MI PMM show a very similar performance in terms of relative bias. In the

simple scenario, the relative bias is very low for m-DAGs T, A, B, D, E, F whereas for m-Dags C, G, I the distortion is higher (-8% to -13%) and greatest for the m-Dags H, J (up to -20%). As the complexity of the DGP increases, so does the distortion regarding the bias. Interestingly, the distortion for the m-Dags T, A, B, D, E, F seems to increase much more than for H, I, J.

The MI methods with interaction terms MI 2Int, MI 3Int, MI 2IntN and MI 3IntN show a very similar distortion across all m-Dags as MI Reg and MI PMM under the simple scenario. There seems to be no difference whether two-way or higher three-way and four-way interactions were included. However, it should be noted that the MI methods 2IntN and MI 3IntN tend to be 2-5 percentage points more distorted. The relative bias remains approximately the same for all m-Dags in the complex 1 scenario and interestingly decreases in the complex 2 scenario. Especially in the m-Dags C, G, H, I, J, this is 7-12 percentage points lower compared to the complex 1 scenario.

Across all three scenarios, the MI CART method produced estimates with comparable levels of bias. This was highest for m-DAGs B, C, G, H, I, and J, ranging from -13% to -27%.

Similarly, the MI RF method also yielded estimates with similar levels of bias across the scenarios and respective m-DAGs, although this was consistently much greater than the bias observed with MI CART. The MI RF method demonstrated the worst performance considering relative bias in simple and complex 1 scenarios compared to all other methods. In the complex 2 scenario, it also performed worst for m-DAGs C, G, H, I, and J. For m-DAGs T, A, B, D, E, and F in the complex 2 scenario, simple parametric methods MI PMM and MI Reg had the highest relative bias, ranging from -28% to -58%.

In the simple scenario, the relative bias for MI Amelia across all m-DAGs was in the range of -7% to -17%. Interestingly, the performance of MI Amelia seems to improve with the increasing complexity of the data, showing relatively low bias in the complex 2 scenario except for m-DAGs F and T.

A single best method concerning the relative bias across all m-DAGs cannot be determined. Nevertheless, the non-MI methods, CC and Ext, performed relatively well for m-DAGs T, A, B, D, E, F, G, with a maximum relative bias of 7%. For m-DAGs H, I, and J, across all scenarios, MI Amelia yielded a relatively low bias.

In figure A.1, the relative bias of the estimated ATE of the presented missing data methods in combination with *TMLE_S2* is displayed. This approach combined with the proposed missing data methods displayed consistent performance in terms of bias across all different scenarios and m-DAGs compared to *TMLE*.

The *LTMLE* in combination with MI methods generally showed a similar pattern regarding bias over the scenarios and m-DAGs (figure A.2) compared to *TMLE* and *TMLE_S2*. However, ATE was consistently overestimated by 1-4 percentage points relative to both *TMLE* and *TMLE_S2*. This characteristic could prove advantageous for those m-DAGs where the ATE is underestimated when using *TMLE* and *TMLE_S2*. A noticeable difference emerges when looking at the non-MI method using CC. There's a clear overestimation when compared to *TMLE* and *TMLE_S2*. This behavior isn't surprising if the estimates based on the full data set (table 4) is considered and compared to CC, particularly with m-DAG T. This overestimation becomes even more significant in the complex scenario 2. How-

ever, what appears intriguing is why the overestimation of the ATE is less pronounced when used in conjunction with MI methods. The overestimation of the ATE for the CC method results in a higher bias, particularly for the recoverable m-DAGs T, A, B, D, and E compared to *TMLE* and *TMLE_S2*. In contrast, for the non-recoverable m-DAGs F, G, H, I, J, this bias significantly decreases.
The *TMLE3* and *CVTMLE3* implementations in combination with missing methods (figures A.3 and A.4) exhibit the same behavior as *LTMLE* and, when used with MI-methods, tend to underestimate the ATE less than when used with CC. For the CC, the overestimation is larger compared to *TMLE* and *TMLE_S2*. Furthermore, it is shown that the relative bias, like the estimate on the complete dataset, is lower for *CVTMLE3* than for TMLE3.

5.1.3 Relative error in model-based standard error

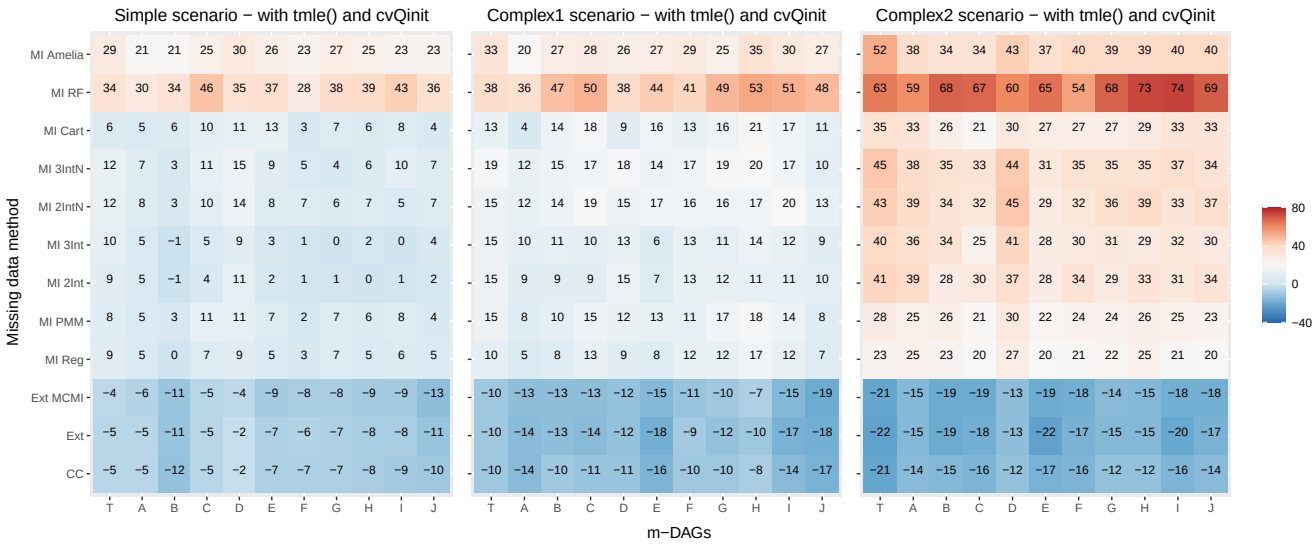


Figure 5: Relative error in ModSE (%) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE* as described in chapter 4.5 in reproduced data.

In the figure 5, the relative error in ModSe of the estimated ATE of the presented missing data methods in combination with *TMLE* is shown.
The relative error in ModSE is very similar for the non-MI methods CC, Ext, and Ext MCMI, with the relative error in ModSE for CC tending to be lower in more complex scenarios. In total, the ModSE is underestimated by CC, Ext, and Ext MCMI (-4% to -21%). The underestimation of the ModSE increases with the complexity of the scenarios.
For all MI methods including MI Amelia, the ModSE is overestimated. The relative error in ModSE strongly increases with the growing complexity of the scenarios. This trend is particularly strong for MI 2Int, MI 3Int, MI 2IntN, and MI 3IntN. MI Amelia and MI RF have the highest relative error in ModSE, with the overestimation of ModSE being significantly much larger for MI RF compared to MI Amelia in the complex 2 scenario. The error in ModSE for the other TMLE implementations in

combination with missing methods can be seen in the figures A.5, A.6, A.7, and A.8 in the appendix.

5.1.4 RMSE

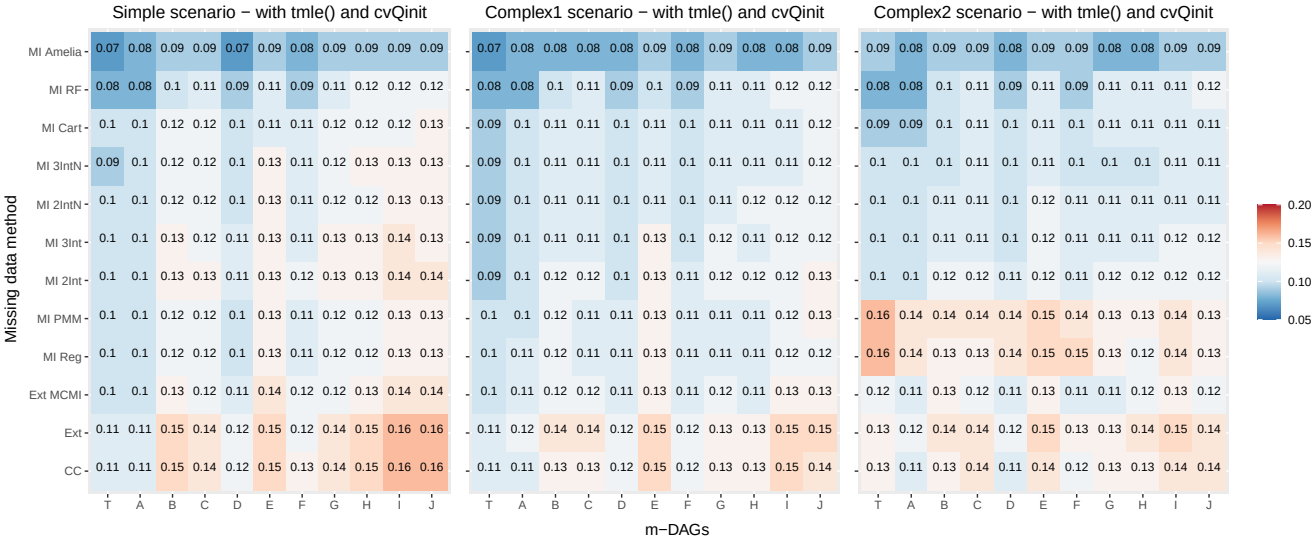


Figure 6: Root mean squared error (RMSE) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE* as described in chapter 4.5 in reproduced data.

In the figure 6, the RMSE of the estimated ATE of the presented missing data methods in combination with *TMLE* is shown.

For the non-MI methods, Complete-Case (CC) and Ext, the root mean square error (RMSE) remained relatively constant as the complexity of the scenarios increased (ranging from 0.11 to 0.16). However, these methods had the highest RMSE for the m-DAGs E, G, H, I, J. On the other hand, Ext MCMI had a lower RMSE across all scenarios (ranging from 0.10 to 0.14).
The RMSE for the parametric MI methods without interactions (MI Reg and MI PMM) increased with the complexity of the scenarios. In the simple scenario and complex 1 scenario, the RMSE was low for m-DAGs T, A, B, C, D, and F (ranging from 0.10 to 0.12) but increased more for m-DAGs G, H, I, J in complex 2 scenario.
The RMSE for the MI methods with interaction terms (MI 2Int, MI 3Int, MI 2IntN, and MI 3IntN) decreased with the complexity of the scenarios and was very homogeneous across all m-DAGs in complex scenarios 1 and 2. In the simple scenario, the RMSE was largest for m-DAGs E, H, I, J and smallest for m-DAGs T and A. The performance regarding the RMSE of the MI methods with CART and RF was almost identical to that of MI 2Int, MI 3Int, MI 2IntN, and MI 3IntN. However, MI RF had a slightly lower RMSE, particularly for m-DAGs T, A, B, D, E. The lowest RMSE across all m-DAGs and scenarios was achieved by MI Amelia (ranging from 0.07 to 0.09). The RMSE for the other TMLE implementations in combination with missing methods can be seen in the figures A.9,

A.10, A.11, and A.12 in the appendix.

5.1.5 Coverage

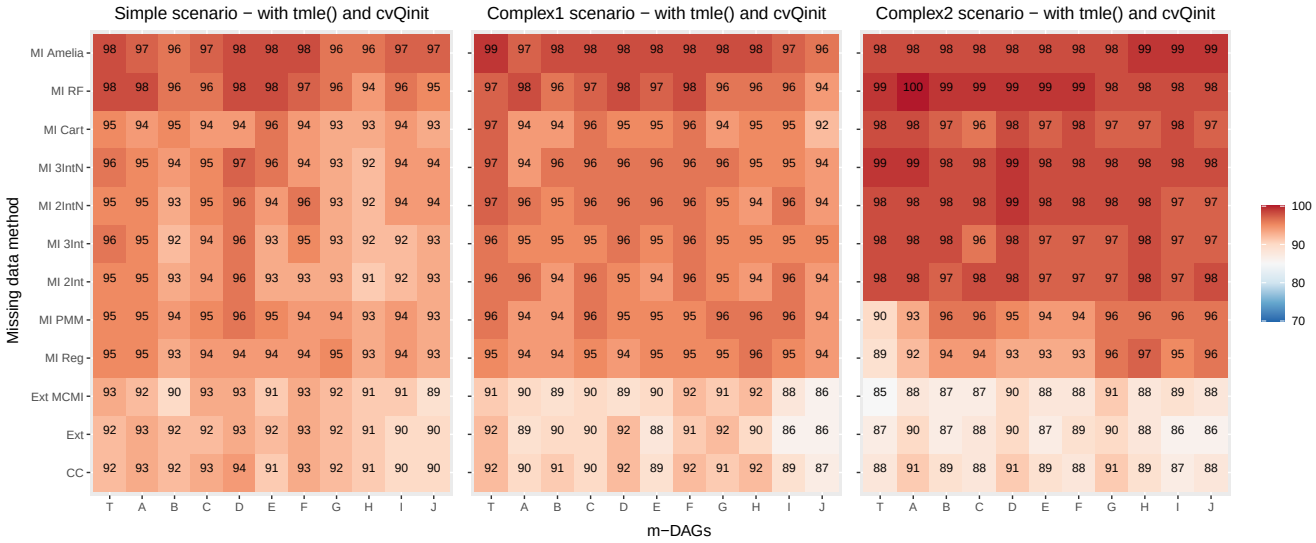


Figure 7: Coverage in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE* as described in chapter 4.5 in reproduced data.

The coverage probabilities of the 95% CI are displayed in figure 7 for the estimated ATE across all m-DAGs and considered sceanrios.

All non-MI methods CC, Ext, and Ext MCMI had very similar coverage probabilities across all m-DAGs and scenarios. However, the coverage probabilities decrease as the complexity of the scenarios increases. Especially for m-DAGs I and J, the coverage probabilities are lower compared to the other m-DAGs.

The coverage probabilities were uniform in behavior for all MI methods across all scenarios. For the m-DAGs H, I, J, the coverage probabilities are lower compared to the other m-DAGs. Moreover, it should be noted that the coverage probabilities increase with the complexity of the scenarios under consideration, and in the complex 2 scenario, they are always above 96% (excluding MI Reg and MI PMM). For MI Amelia and MI CART, the coverage probabilities in the more complex scenarios are very high (94%-100%) for all m-DAGs.

For MI Reg and MI PMM under simple and complex 1 scenarios, the coverage probabilities were relatively constant (93%-96%) and also tend to increase with the complexity of the scenarios, although these are significantly lower under m-DAG T (89% and 90%). Further, in tables B.11, B.12, and B.13, one can observe all performance measures for the *TMLE* implementation in relation to the 11 assessed missingness mechanisms across all scenarios. In the results, it can be observed that for MI using Amelia, the empSE is the lowest across all m-dags and scenarios. Although MI Amelia, along with MI RF, tends to overestimate the ModSE to some extent, this overestimation is not problematic when considering the length of the CI and comparing it to other MI methods. This is because the

non-MI methods, CC and Ext, underestimate the ModSE and have a similar CI length. Overall, it can be concluded that MI Amelia has a high coverage mainly due to its high precision (low empSE and RMSE). Since MI Amelia also exhibits no substantial bias, the bias-eliminated coverage is only slightly higher. The RMSE for the other TMLE implementations in combination with missing methods can be seen in the figures A.13, A.14, A.15, and A.16 in the appendix.

5.2 Results of positivity violation data

5.2.1 Pre-analysis

	Estimated mean	Bias	Relative Bias (%)	empSE	RMSE	ModSE	Error in ModSE (%)	Coverage	Bias eliminated coverage	Mean CI length	Proportional CI length
Positivity violation 1 and complex 1 scenario											
TMLE	0.17	-0.03	-12.78	0.12	0.12	0.10	-16.20	0.88	0.89	0.39	0.99
TMLE_S2	0.17	-0.03	-14.08	0.12	0.12	0.10	-16.89	0.87	0.88	0.38	0.95
LTMLE	0.13	-0.07	-35.05	0.12	0.14	0.10	-17.40	0.81	0.87	0.39	1.00
LTMLE_ic	0.13	-0.07	-34.95	0.12	0.14	0.10	-17.66	0.80	0.88	0.39	1.00
CVTMLE3	0.19	-0.01	-7.03	0.12	0.12	0.08	-31.35	0.82	0.80	0.33	0.83
TMLE3	0.19	-0.01	-6.41	0.12	0.12	0.08	-36.00	0.78	0.79	0.30	0.77
Positivity violation 1 and complex 2 scenario											
TMLE	0.18	-0.02	-10.66	0.12	0.13	0.10	-17.89	0.88	0.88	0.39	0.95
TMLE_S2	0.18	-0.02	-11.56	0.12	0.12	0.10	-16.85	0.88	0.89	0.38	0.93
LTMLE	0.12	-0.08	-39.29	0.13	0.15	0.11	-19.93	0.80	0.87	0.41	1.00
LTMLE_ic	0.12	-0.08	-39.31	0.13	0.15	0.11	-19.93	0.79	0.88	0.41	1.00
CVTMLE3	0.20	0.00	1.30	0.13	0.13	0.09	-34.38	0.80	0.80	0.33	0.81
TMLE3	0.21	0.01	3.39	0.13	0.13	0.08	-38.60	0.76	0.76	0.31	0.76
Positivity violation 2 and complex 1 scenario											
TMLE	0.17	-0.03	-16.87	0.12	0.12	0.08	-32.08	0.76	0.80	0.30	1.00
TMLE_S2	0.17	-0.03	-16.89	0.11	0.12	0.08	-32.79	0.76	0.79	0.29	0.97
LTMLE	0.08	-0.12	-59.52	0.12	0.17	0.08	-36.19	0.54	0.77	0.29	0.99
LTMLE_ic	0.08	-0.12	-60.00	0.12	0.17	0.08	-35.18	0.54	0.78	0.29	0.98
CVTMLE3	0.19	-0.01	-6.60	0.14	0.14	0.07	-50.98	0.63	0.65	0.26	0.88
TMLE3	0.19	-0.01	-3.92	0.14	0.14	0.06	-54.77	0.61	0.61	0.24	0.81
Positivity violation 2 and complex 2 scenario											
TMLE	0.16	-0.04	-17.88	0.11	0.12	0.08	-29.53	0.79	0.82	0.30	0.97
TMLE_S2	0.16	-0.04	-18.10	0.11	0.11	0.08	-30.08	0.79	0.82	0.29	0.94
LTMLE	0.06	-0.14	-70.98	0.12	0.19	0.08	-35.11	0.51	0.77	0.31	1.00
LTMLE_ic	0.06	-0.14	-70.63	0.12	0.19	0.08	-35.11	0.50	0.78	0.31	1.00
CVTMLE3	0.23	0.03	13.32	0.14	0.15	0.07	-51.74	0.61	0.63	0.27	0.87
TMLE3	0.23	0.03	14.82	0.15	0.15	0.06	-56.43	0.58	0.59	0.25	0.81

Table 5: Results for various TMLE implementations on complete positivity violation data across all scenarios

Table 5 presents the results for different TMLE implementations across all scenarios for the positivity data. In scenarios with a 30% positivity violation, the implementations *CVTMLE3* and *TMLE3* exhibited relatively low bias in both scenarios. However, the relative bias was significantly higher (-11% to -14%) for *TMLE* and *TMLE_S2*, and even worse for *LTMLE* and *LTMLE_ic* (-35% to -39%). The empirical SE and RMSE were higher for *LTMLE* and *LTMLE_ic* due to their higher bias. On the

other hand, the error in ModSE was greatly underestimated for *CVTMLE3* and *TMLE3* (-31% to -38%), leading to narrower CIs and consequently worse coverage. In the case of *LTMLE* and *LTMLE_ic*, despite their longer CI lengths, the poorer coverage can be attributed to their high bias. The error in ModSE is approximately the same for *TMLE*, *TMLE_S2*, *LTMLE*, and *LTMLE_ic*. It is worth noting that the ModSE is consistently underestimated for all implementations.

Under the scenarios with a 40% positivity violation, the bias for *TMLE* and *TMLE_S2* decreased by 17% to 18%. For *CVTMLE3* and *TMLE3*, the bias for the ATE only increased under complex 2 scenario (14%). For *LTMLE* and *LTMLE_ic*, the relative bias increased to -60% in complex 1 scenario and -71% in complex 2 scenario. Compared to positivity violation 1, the empSE increased for *CVTMLE3* and *TMLE3*, resulting in a significantly greater underestimation of the ModSE and consequently leading to poor coverage (0.58 to 0.63). Despite having long CIs, *LTMLE* and *LTMLE_ic* resulted in poor coverage due to bias, as evident from the bias eliminated coverage. In terms of coverage and RMSE, *TMLE* and *TMLE_S2* achieved the best performance but underestimated the ATE more than *CVTMLE3* and *TMLE3*. Among all scenarios and positivity violations, *CVTMLE3* exhibited a lower underestimation of the ModSE compared to *TMLE3*.

5.2.2 Bias

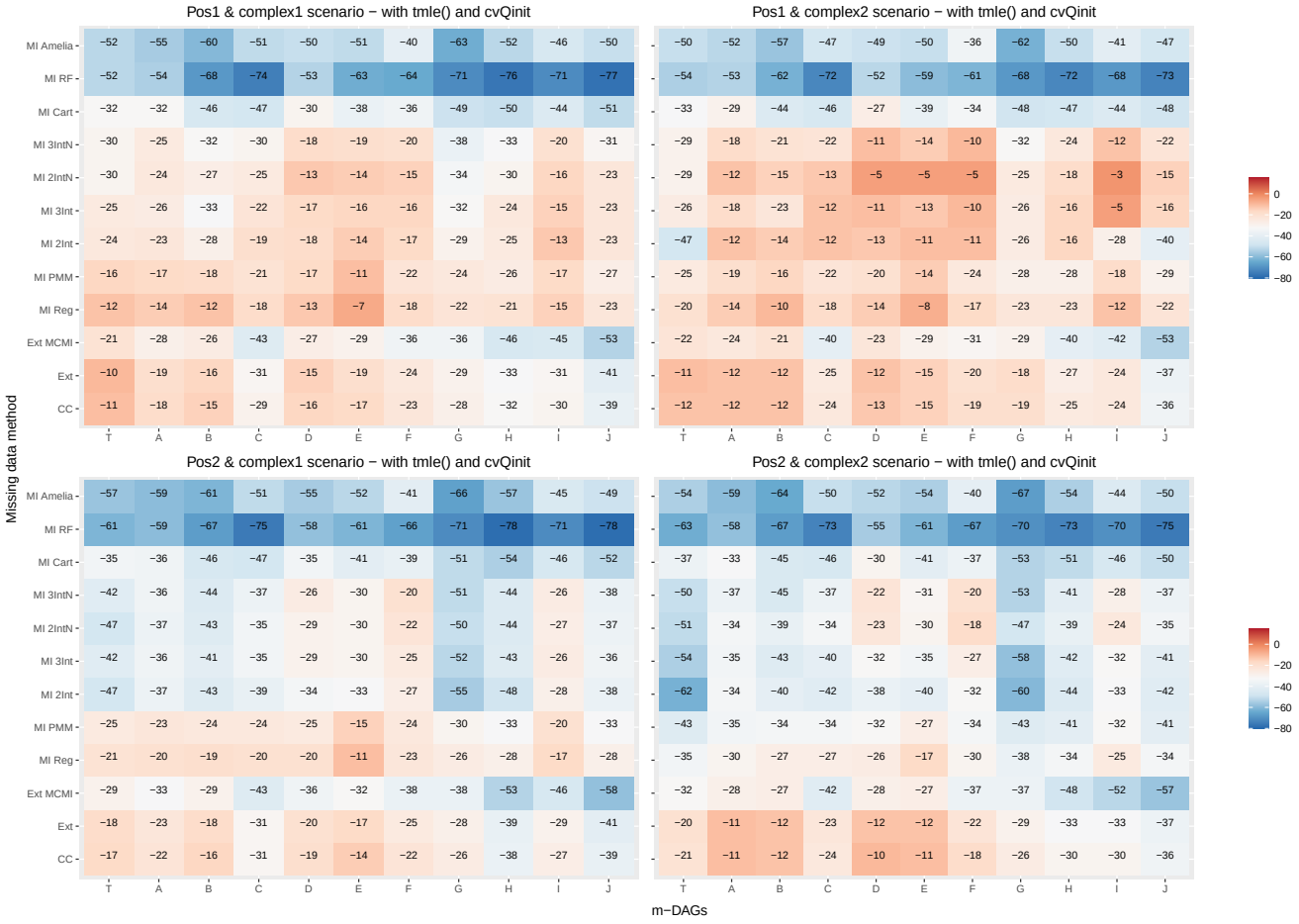


Figure 8: Relative bias (%) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE* as described in chapter 4.5 in positivity data.

In figure 8, the relative bias of the estimated ATE of the presented missing data methods in combination with *TMLE* in the positivity violation data is shown.

Taking the bias on the complete data from table 5 into account, it can be stated that the non-MI methods CC and Ext have less bias for the m-DAGs T, A, B, D, E, except for the complex 1 scenario with a positivity violation of 30%. The bias is larger for the non-recoverable m-DAGs H and J. All MI methods exhibit higher bias in both scenarios with a 40% positivity violation. MI Amelia, MI RF, and MI CART have the largest bias across all scenarios and m-DAGs.

5.2.3 Relative error in ModSe

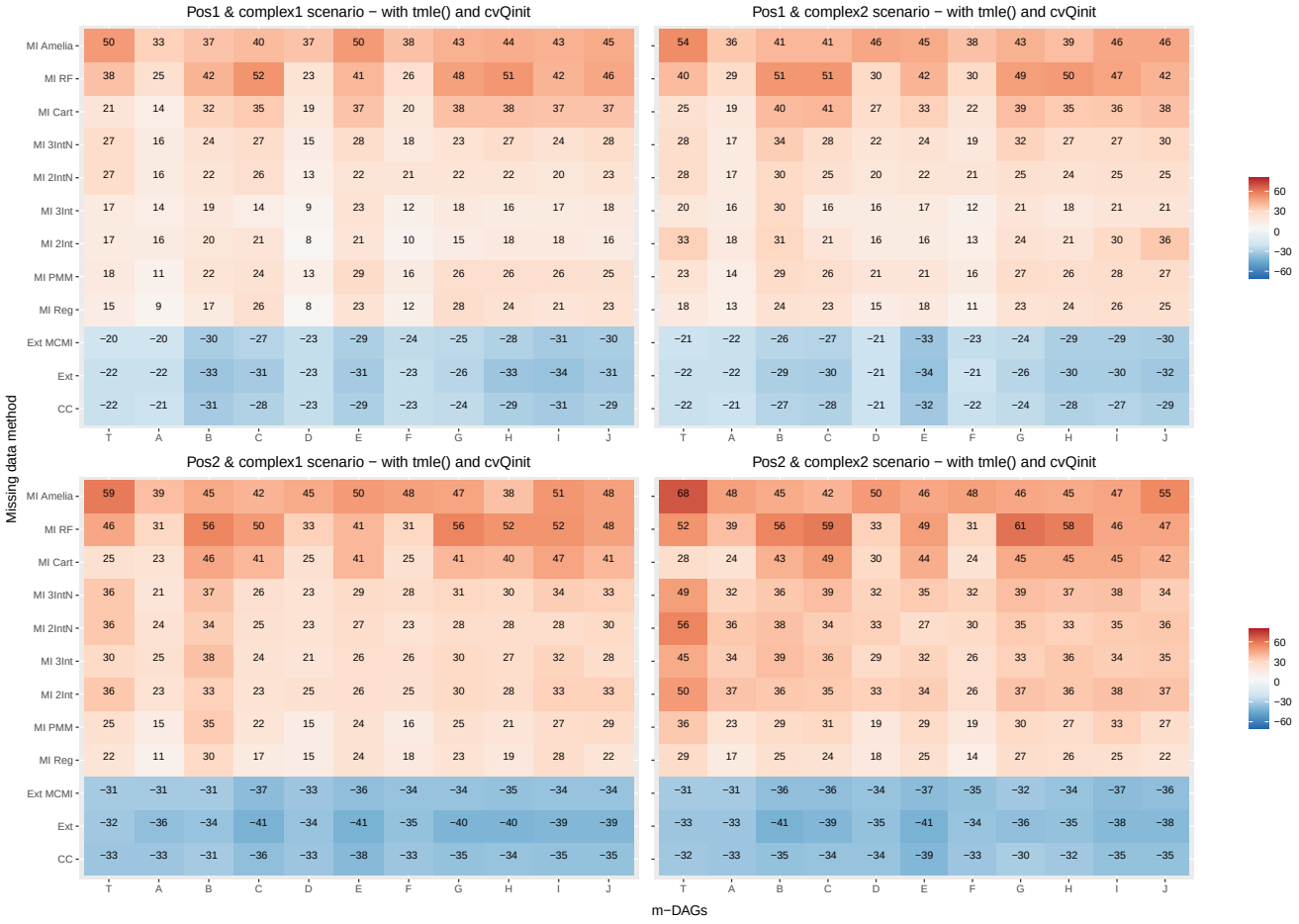


Figure 9: Relative error in ModSE (%) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE* as described in chapter 4.5 in positivity data.

In the figure 9, the relative error in ModSe of the estimated ATE of the presented missing data methods in combination with *TMLE* is shown.

All non-MI methods underestimated the ModSEs similarly across all m-DAGs and scenarios. The underestimation increases with the positivity violation. In contrast, MI methods tended to overestimate the ModSE, and this overestimation increases as the positivity violation increases. The nonparametric MI methods, MI Amelia, MI RF, and MI CART significantly overestimate the ModSE. For m-DAGs C, H, I, and J, the error in ModSE is the highest across all scenarios and positivity violations across all missing data methods.

5.2.4 RMSE

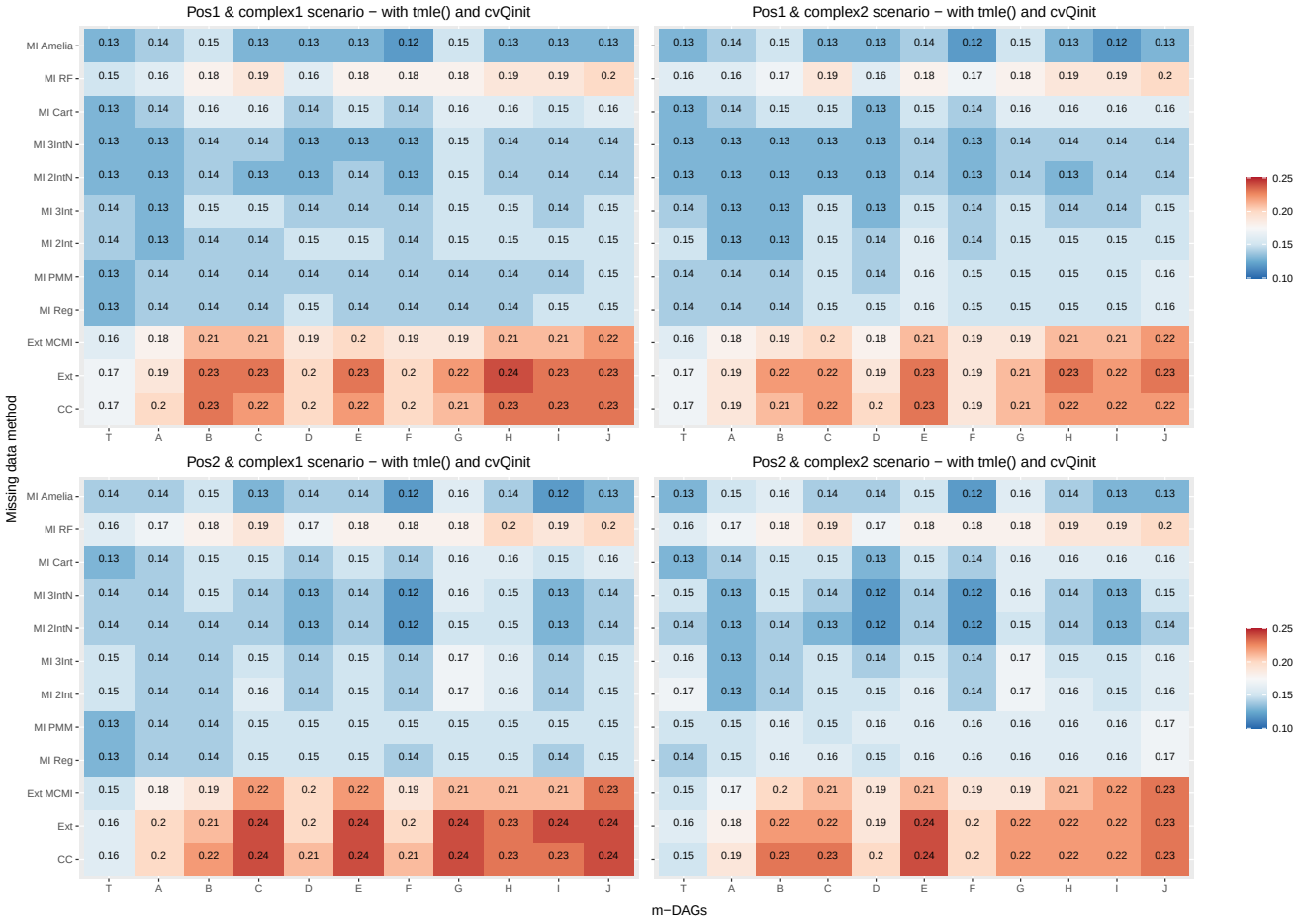


Figure 10: Root mean squared error (RMSE) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE* as described in chapter 4.5 in positivity data.

Figure 10 shows the RMSE of the estimated ATE of the presented missing data methods in combination with *TMLE*. For all non-MI methods, the RMSE is significantly worse compared to the MI methods. Among them, Ext MCMI tends to have a lower RMSE compared to CC and Ext. The RMSE is particularly poor for the m-DAGs G, H, I, J. Among the MI methods, MI Amelia stands out again, consistently demonstrating a low RMSE for all m-Dags across all scenarios and positivity violations.

5.2.5 Coverage

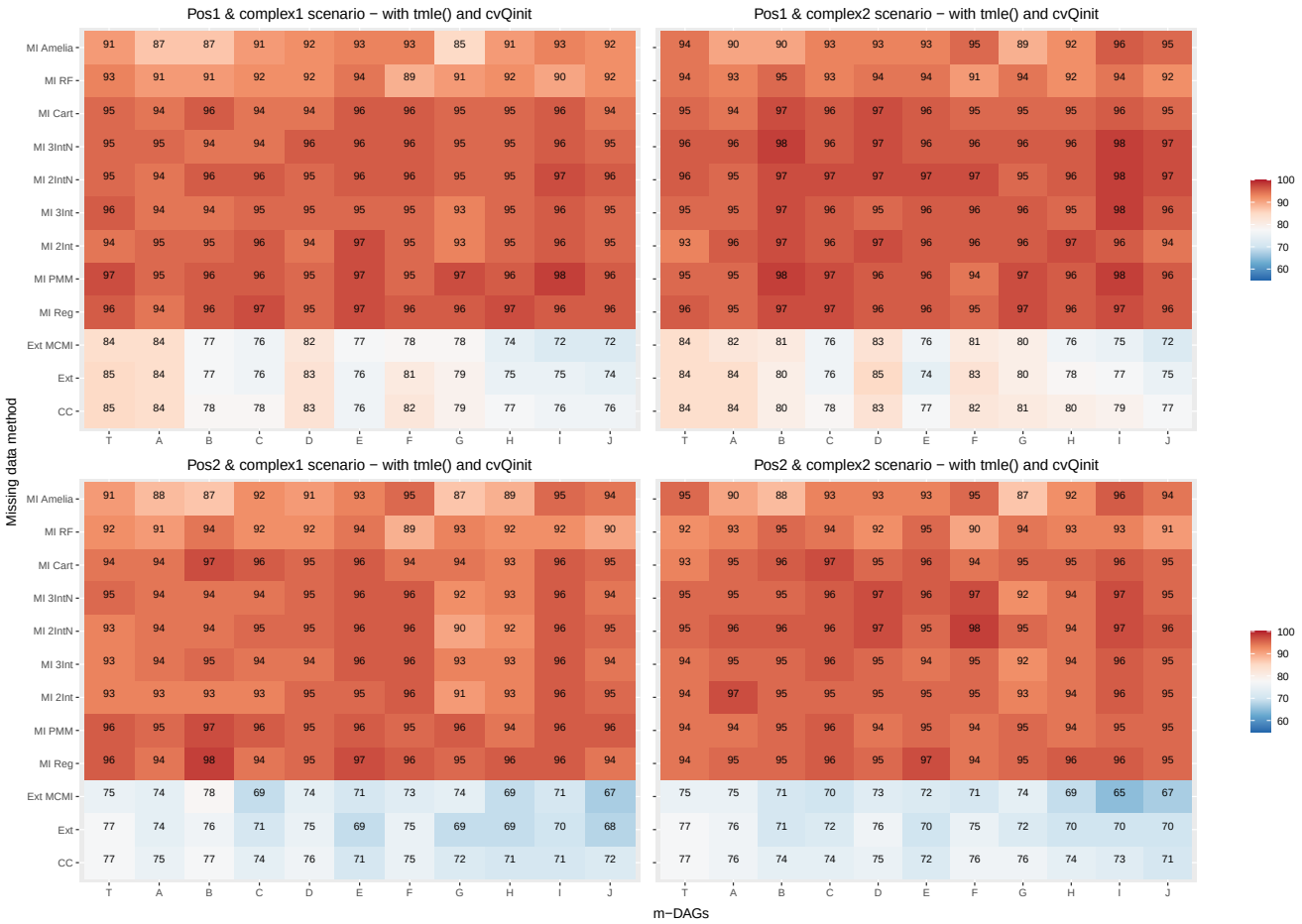


Figure 11: Coverage in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE* as described in chapter 4.5 in positivity data.

In Figure 11, it can be observed that underestimating the ModSE for the non-MI methods leads to significantly lower coverage. The coverage decreases with higher positivity violation. Among the four scenarios, MI PMM and MI Reg have the highest coverage. MI Amelia performs significantly worse in terms of coverage compared to the other MI methods. The reason for this lies in the significant bias combined with a relatively short CI.

5.3 Results of modified data

In the subsequent chapter, estimation of the ATE is solely performed using *TMLE_S2*, owing to its shorter runtime advantage.

5.3.1 Pre-analysis

	Estimated mean	Bias	Relative Bias (%)	empSE	RMSE	ModSE	Error in ModSE (%)	Coverage	Bias eliminated coverage	Mean CI length	Proportional CI length
Simple scenario											
TMLE	0.20	-0.00	-1.03	0.07	0.07	0.07	-5.76	0.93	0.93	0.27	0.90
TMLE_S2	0.20	-0.00	-0.85	0.07	0.07	0.07	-6.87	0.92	0.92	0.26	0.88
LTMLE	0.21	0.01	4.64	0.07	0.08	0.07	-4.43	0.94	0.93	0.28	0.93
LTMLE_ic	0.21	0.01	4.37	0.07	0.08	0.07	-4.08	0.94	0.94	0.28	0.93
CVTMLE3	0.20	-0.00	-0.73	0.08	0.08	0.08	0.04	0.95	0.95	0.30	1.00
TMLE3	0.20	0.00	0.08	0.08	0.08	0.07	-11.75	0.91	0.91	0.26	0.87
Complex 1 scenario											
TMLE	0.21	0.01	4.40	0.11	0.11	0.08	-28.80	0.80	0.80	0.30	0.93
TMLE_S2	0.21	0.01	3.94	0.11	0.11	0.08	-30.24	0.79	0.79	0.29	0.89
LTMLE	0.25	0.05	23.12	0.11	0.12	0.09	-24.75	0.79	0.83	0.33	1.00
LTMLE_ic	0.25	0.05	23.23	0.11	0.12	0.09	-24.50	0.80	0.84	0.33	1.00
CVTMLE3	0.21	0.01	5.07	0.12	0.12	0.08	-32.36	0.80	0.79	0.32	0.98
TMLE3	0.21	0.01	6.50	0.12	0.12	0.07	-45.26	0.68	0.68	0.26	0.79
Complex 2 scenario											
TMLE	0.24	0.04	18.60	0.16	0.17	0.09	-45.44	0.64	0.66	0.33	0.85
TMLE_S2	0.23	0.03	17.02	0.16	0.16	0.08	-48.16	0.63	0.64	0.30	0.79
LTMLE	0.39	0.19	95.68	0.22	0.29	0.10	-53.59	0.46	0.59	0.38	1.00
LTMLE_ic	0.39	0.19	97.07	0.23	0.30	0.10	-56.15	0.43	0.56	0.38	0.99
CVTMLE3	0.25	0.05	25.50	0.20	0.20	0.08	-57.55	0.58	0.59	0.32	0.84
TMLE3	0.25	0.05	25.90	0.20	0.21	0.06	-67.88	0.46	0.46	0.25	0.65

Table 6: Results for various TMLE implementations on complete modified data across all scenarios

Table 6 presents the performance of different TMLE implementations across all scenarios for the modified data. The results for the simple scenario are very similar compared to those obtained from the reproduced data. However, in general, the error in ModSE for all TMLE implementations is approximately twice as large. An exception to this is *CVTMLE3*, which estimates the ModSE accurately. Nonetheless, the estimates for the more complex scenarios are less precise for all TMLE implementations, as indicated by larger empSE and RMSE values. Among these, *TMLE3* exhibits the greatest underestimation of ModSE. For *TMLE*, *TMLE_S2*, *CVTMLE3*, and *TMLE3*, the poorer coverage is again due to the significant underestimation of ModSE, while for *LTMLE* and *LTMLE_ic*, the large bias also affects the coverage. In terms of coverage, RMSE, and bias, *TMLE* and *TMLE_S2* achieved the best performance across all scenarios. In the complex 2 scenario, *LTMLE* and *LTMLE_ic* exhibited extremely high relative biases of 96% and 97%.

5.3.2 Bias

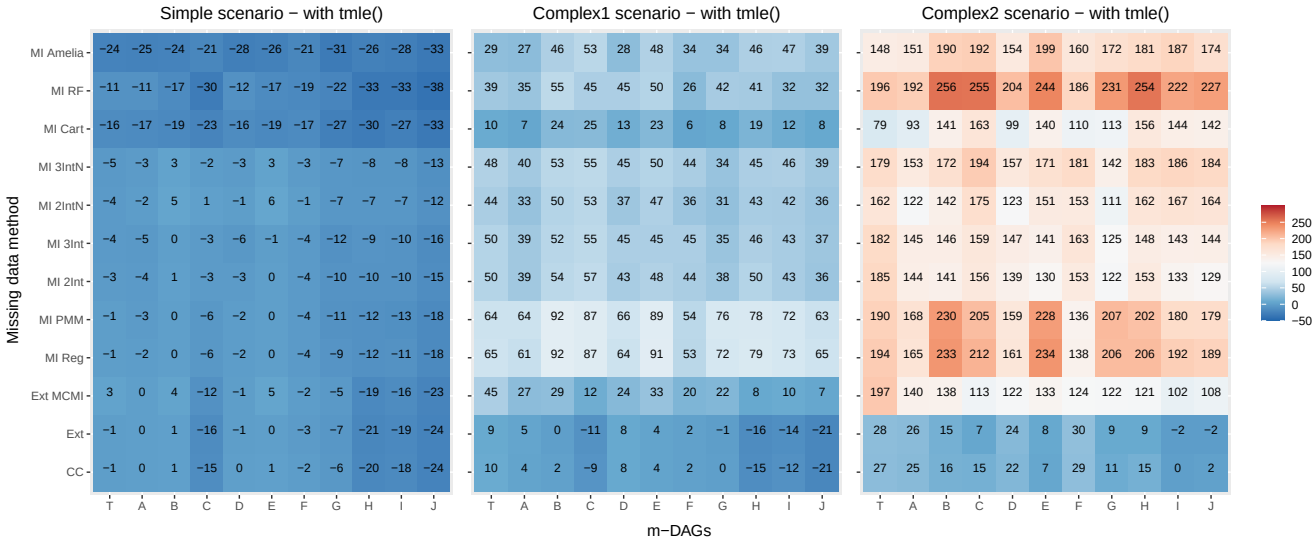


Figure 12: Relative bias (%) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE_S2* as described in chapter 4.5 in modified data.

In figure 12, the relative bias of the estimated ATE of the presented missing data methods in combination with *TMLE* in the modified data is shown.
Taking into account the bias on the complete data from table 6, it can be stated that the non-MI methods CC and Ext are unbiased for the m-DAGs T, A, B, D, E, F, G except for the complex 2 scenario. For the complex 2 scenario m-DAGs B, C, E, G, H, I, J seem to have low relative bias. The bias is larger for the non-recoverable m-DAGs H and J. Ext MCMI exhibits a significantly larger bias across all m-DAGs compared to CC and Ext in the complex 1 and complex 2 scenarios. All MI methods exhibit higher in both complex scenarios. MI Amelia and MI RF have the largest bias across all scenarios and m-DAGs. For the simple scenario 1, all methods exhibit similar behavior in terms of relative bias, except for MI Amelia, MI CART, and MI RF. However, in the complex 1 scenario and complex 2 scenario, the MI methods show significantly higher bias, particularly in complex 2 scenario
.

5.3.3 Relative error in ModSe

For the non-MI methods, the underestimation of ModSE increases with the complexity of the scenarios (figure 13). However, in all scenarios, it is greater for the m-DAGs H, I, J. In complex 2 scenario, the underestimation is relatively homogeneous across all m-DAGs. Among the MI methods, MI CART stands out, as the overestimation of ModSE is relatively low across all scenarios and m-DAGs. It should be noted that all MI methods overestimate ModSE in the complex scenarios.

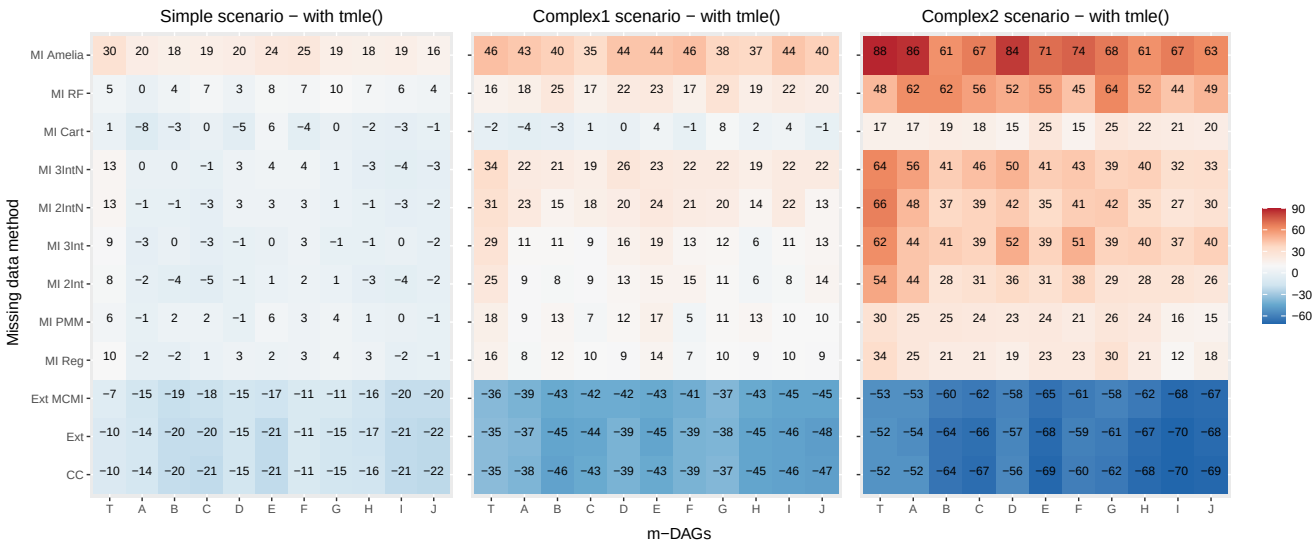


Figure 13: Relative error in ModSE (%) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE_S2* as described in chapter 4.5 in modified data.

5.3.4 RMSE

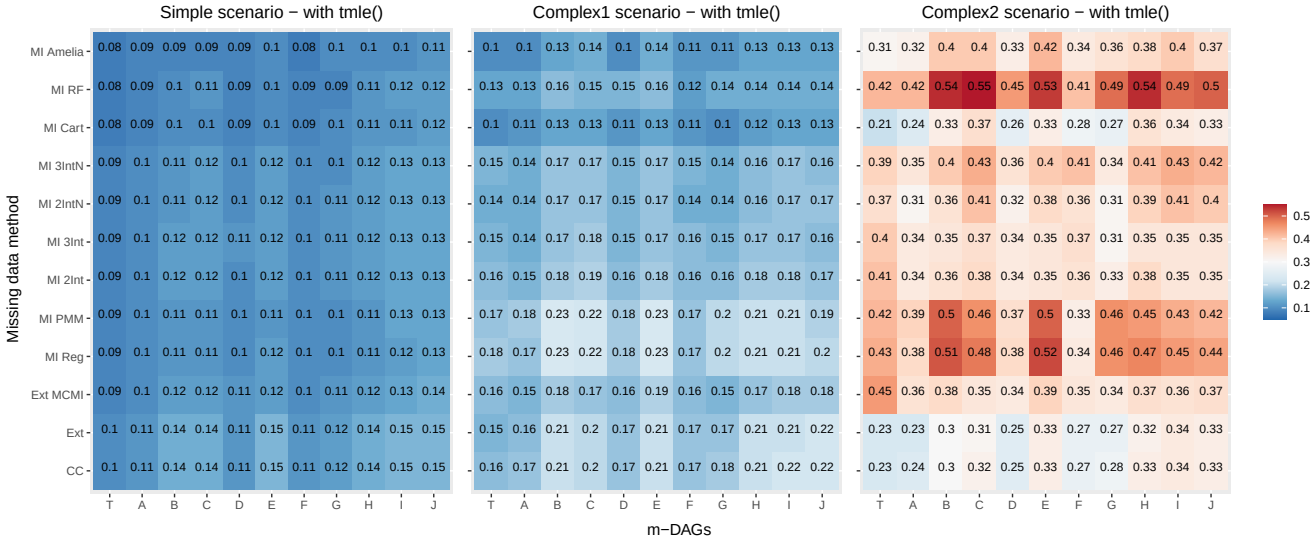


Figure 14: Root mean squared error (RMSE) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE_S2* as described in chapter 4.5 in modified data.

For the complex 1 and complex 2 scenarios, MI Amelia and MI CART have the lowest RMSE compared to the other missing methods (figure 14). In general, the non-MI methods have a higher RMSE. However, this changes in the complex 2 scenario, which is due to the fact that the bias for the MI methods was much higher here.

5.3.5 Coverage

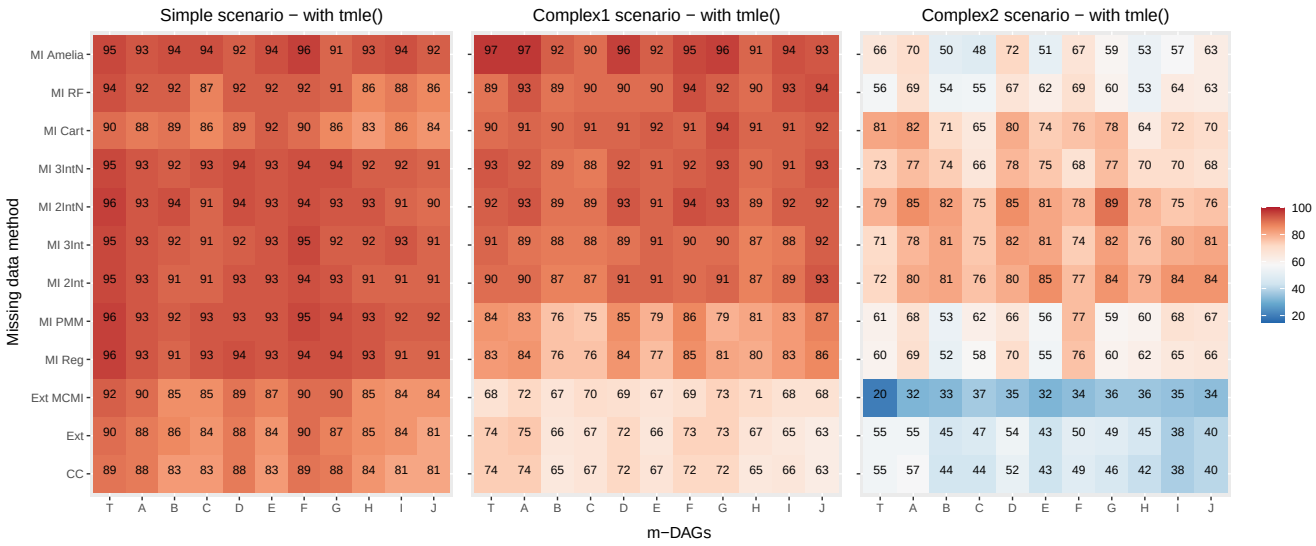


Figure 15: Coverage in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE_S2* as described in chapter 4.5 in modified data.

The coverage is relatively high for the MI methods in the simple scenario and decreases slightly in the complex 1 scenario (figure 15). In the complex 2 scenario, it is significantly lower, which can be attributed to the high bias. The non-MI methods exhibit very poor coverage, especially in the complex 2 scenario. This is primarily due to a significant underestimation of the ModSE and to a lesser extent, the bias.

5.4 Results of copula 1 data

5.4.1 Pre-analysis

	Estimated mean	Bias	Relative Bias (%)	empSE	RMSE	ModSE	Error in ModSE (%)	Coverage	Bias eliminated coverage	Mean CI length	Proportional CI length
Simple scenario											
TMLE	0.20	-0.00	-1.76	0.08	0.08	0.08	-8.91	0.90	0.90	0.29	0.91
TMLE_S2	0.20	-0.00	-1.82	0.08	0.08	0.07	-9.41	0.90	0.90	0.28	0.89
CVTMLE3	0.20	-0.00	-1.59	0.09	0.09	0.08	-6.64	0.93	0.92	0.31	1.00
TMLE3	0.20	-0.00	-0.35	0.09	0.09	0.07	-18.84	0.86	0.86	0.27	0.86
Complex 1 scenario											
TMLE	0.21	0.01	3.52	0.10	0.10	0.07	-30.83	0.80	0.80	0.27	0.90
TMLE_S2	0.21	0.01	2.95	0.10	0.10	0.07	-32.68	0.78	0.78	0.26	0.86
CVTMLE3	0.22	0.02	11.00	0.11	0.12	0.08	-32.02	0.80	0.81	0.30	1.00
TMLE3	0.24	0.04	18.64	0.12	0.12	0.06	-46.72	0.67	0.69	0.24	0.79
Complex 2 scenario											
TMLE	0.23	0.03	12.59	0.14	0.14	0.08	-44.72	0.67	0.66	0.28	0.96
TMLE_S2	0.22	0.02	10.22	0.13	0.14	0.07	-45.69	0.67	0.67	0.27	0.91
CVTMLE3	0.25	0.05	27.41	0.17	0.18	0.08	-55.21	0.59	0.61	0.30	1.00
TMLE3	0.28	0.08	38.86	0.19	0.20	0.06	-68.06	0.45	0.49	0.23	0.77

Table 7: Results for various TMLE implementations on complete copula 1 data across all scenarios

Table 7 presents the results of the performance of different TMLE implementations across all scenarios for the copula 1 data. The implementations with the **ltmle** package were not further considered for the copula 1, copula 2, and copula 3 data due to significant increases in runtime when using continuous confounders. The results for the simple scenario are very similar compared to those obtained from the modified data. However, in general, the error in ModSE for all TMLE implementations is larger, resulting in worse coverage. As the complexity of the scenarios increases, the overestimation of the ATE also increases, particularly in complex 2 scenario for *CVTMLE3* and *TMLE3*, with a relative bias of 27% and 39%, respectively, while it remains moderate at 13% and 10% for *TMLE* and *TMLE_S2*. Additionally, the empSE and RMSE also increase, particularly for *CVTMLE3* and *TMLE3*. Moreover, the ModSE is significantly underestimated, especially for *TMLE3*. Once again, in terms of coverage, RMSE, bias, and estimation of ModSE, *TMLE* and *TMLE_S2* achieved the best performance across all scenarios, with similar performance as observed previously. Finally *CVTMLE3* has the widest CI in all scenarios.

5.4.2 Bias

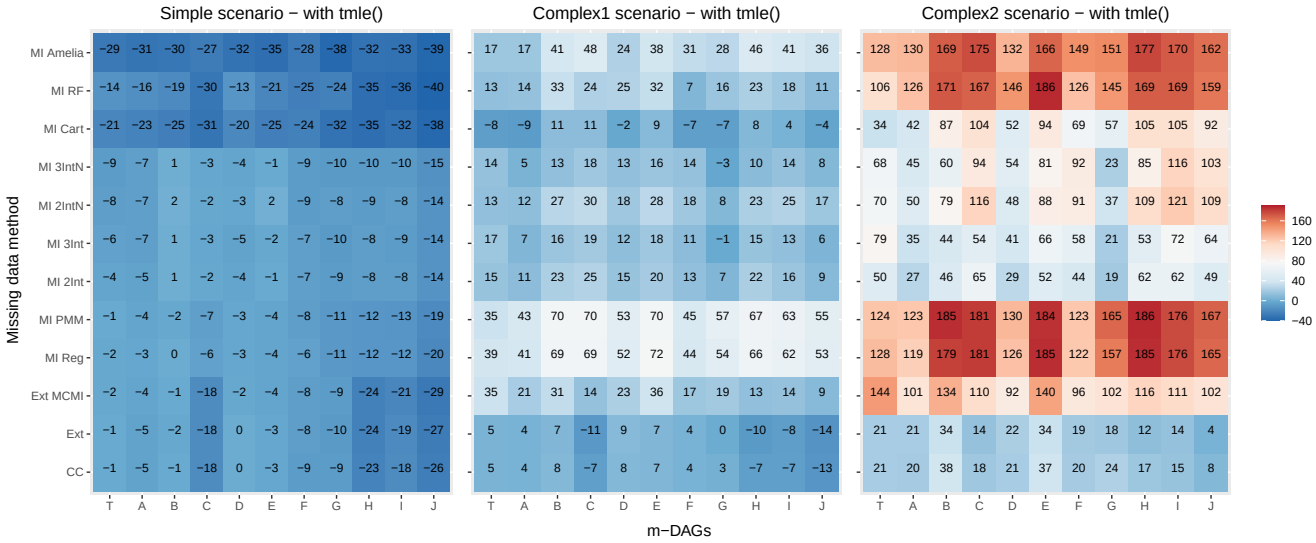


Figure 16: Relative bias (%) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE_S2* as described in chapter 4.5 in copula 1 data.

For the non-MI methods CC and Ext, the bias in the complex 1 scenario was low for m-DAGs T, A, B, D, E, F, G, but also moderate for the other non-recoverable m-DAGs H, J, and C (see figure 16). In the complex 2 scenario, the relative bias was significantly larger, but it was lower for the non-recoverable m-DAGs F, G, H, I, J compared to the recoverable m-DAGs. The simple scenario reflects the expected results, with lower bias for m-DAGs T, A, B, D, E. However, it can be concluded that CC and Ext outperform the MI methods in terms of bias in the complex scenarios.

5.4.3 Relative error in ModSe

For the non-MI methods, the underestimation of ModSE increases with the complexity of the scenarios (figure 17) similar to modified data. However, in the simpler scenarios, it is smaller for the m-DAGs T, A, F. In complex 2 scenario, the underestimation is relatively homogeneous across all m-DAGs. Among the MI methods, MI CART stands out, as the overestimation of ModSE is very low across all m-DAGs and also in complex 2 scenario it is superior compared to the other methods. It should be noted that all MI methods overestimate the ModSE in the complex scenarios. Ext MCMI exhibits a significantly larger bias across all m-DAGs compared to CC and Ext in the complex 1 and complex 2 scenarios.

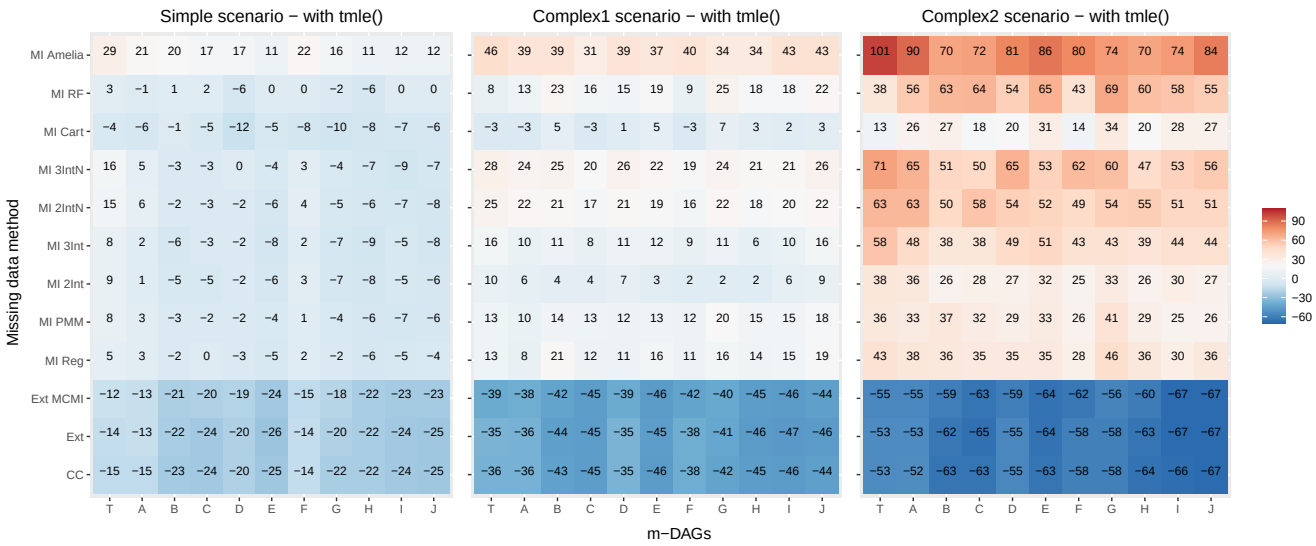


Figure 17: Relative error in ModSE (%) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE_S2* as described in chapter 4.5 in copula 1 data.

5.4.4 RMSE

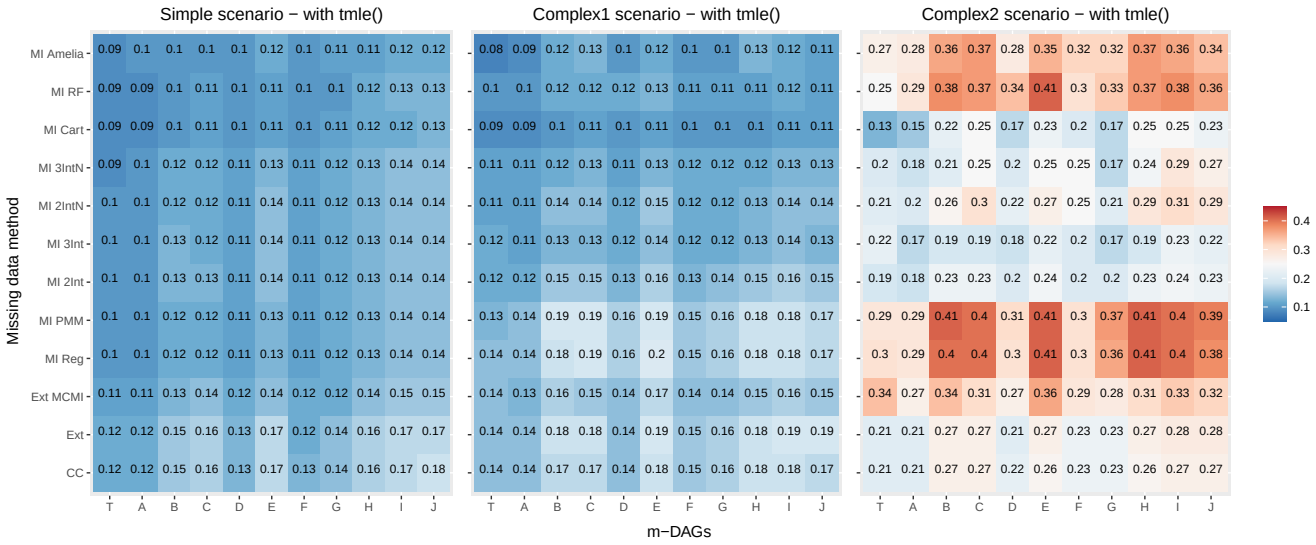


Figure 18: Root mean squared error (RMSE) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE_S2* as described in chapter 4.5 in copula 1 data.

For the simple and complex 1 scenarios, MI Amelia and MI CART have the lowest RMSE compared to the other missing methods (figure 18). And in complex 2 scenario MI Cart and MI 3Int were superior across all m-DAGs. In general, the non-MI methods have a higher RMSE in the simpler scenarios, whereas it changes in the complex 2 scenario, which is due to the fact that the bias for the MI methods was much higher.

5.4.5 Coverage

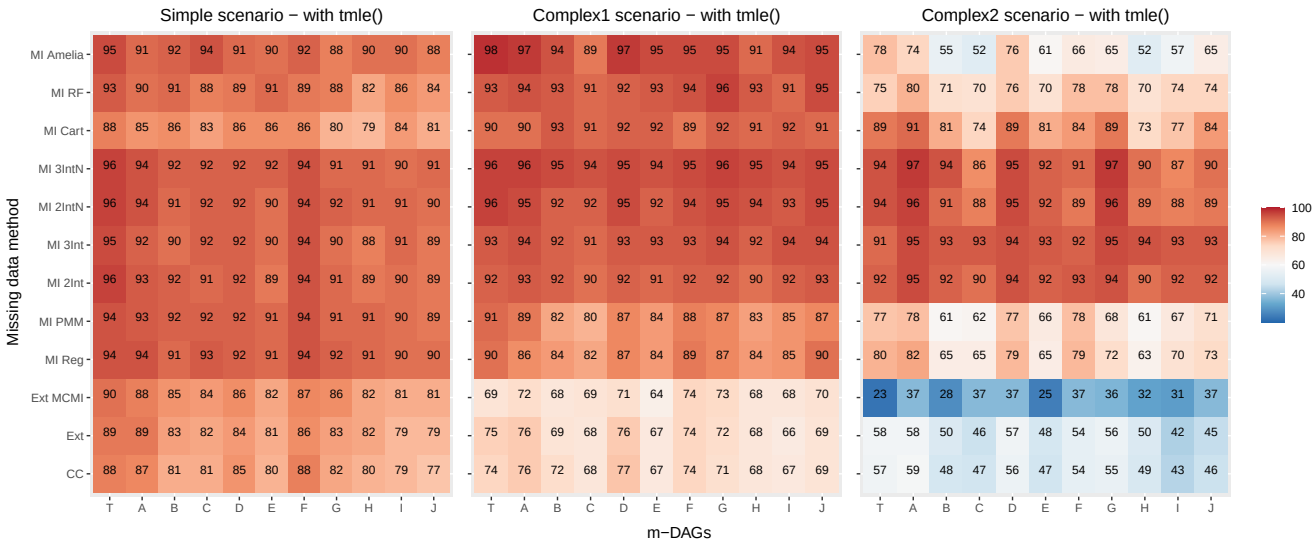


Figure 19: Coverage in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE_S2* as described in chapter 4.5 in copula 1 data.

The coverage is relatively high for the MI methods in the complex 1 scenario (figure 19). In the complex 2 scenario, it is significantly lower, which can be attributed to the high bias. The non-MI methods togehter with MI Amelia exhibit very poor coverage, especially in the complex 2 scenario. This is primarily due to a significant underestimation of the ModSE and to a lesser extent, the bias.

5.5 Results of copula 2 data

5.5.1 Pre-analysis

	Estimated mean	Bias	Relative Bias (%)	empSE	RMSE	ModSE	Error in ModSE (%)	Coverage	Bias eliminated coverage	Mean CI length	Proportional CI length
Simple scenario											
TMLE	0.20	-0.00	-0.09	0.09	0.09	0.08	-10.30	0.91	0.91	0.30	0.94
TMLE_S2	0.20	-0.00	-0.54	0.09	0.09	0.08	-11.77	0.91	0.91	0.30	0.92
CVTMLE3	0.20	0.00	0.33	0.09	0.09	0.08	-8.77	0.92	0.92	0.32	1.00
TMLE3	0.20	0.00	1.30	0.09	0.09	0.07	-19.66	0.87	0.87	0.29	0.89
Complex 1 scenario											
TMLE	0.20	-0.00	-2.24	0.11	0.11	0.08	-31.19	0.79	0.79	0.29	0.93
TMLE_S2	0.19	-0.01	-4.70	0.10	0.10	0.07	-33.40	0.78	0.77	0.26	0.86
CVTMLE3	0.21	0.01	5.35	0.12	0.12	0.08	-34.32	0.78	0.79	0.31	1.00
TMLE3	0.22	0.02	11.32	0.12	0.13	0.07	-46.61	0.69	0.69	0.25	0.82
Complex 2 scenario											
TMLE	0.19	-0.01	-6.76	0.15	0.15	0.08	-48.23	0.64	0.64	0.28	0.95
TMLE_S2	0.18	-0.02	-8.92	0.14	0.14	0.07	-51.01	0.63	0.63	0.26	0.87
CVTMLE3	0.17	-0.03	-13.06	0.18	0.18	0.08	-57.21	0.60	0.61	0.30	1.00
TMLE3	0.21	0.01	2.51	0.19	0.19	0.06	-66.91	0.49	0.48	0.24	0.80

Table 8: Results for various TMLE implementations on complete copula 2 data across all scenarios

Table 8 contains the results of the performance of different TMLE implementations across all scenarios for the copula 2 data. The results for the simple scenario are very similar to those obtained from the Copula 1 data for all performance measures. Similarly, for the complex 1 and complex 2 scenarios, empSE, RMSE, coverage, CI length, and error in ModSE are comparable. However, a difference can be observed in the relative bias, which is lower in both complex scenarios for Copula 2 data. In particular, *CVTMLE3* and *TMLE3* exhibit significantly lower bias. In complex 2 sceanrio the relative bias was also lower for *TMLE* and *TMLE_S2*. Overall, the coverage for *TMLE* and *TMLE_S2* is also better than that for *CVTMLE3* and *TMLE3*, although *CVTMLE3* has the widest CI in all scenarios.

5.5.2 Bias

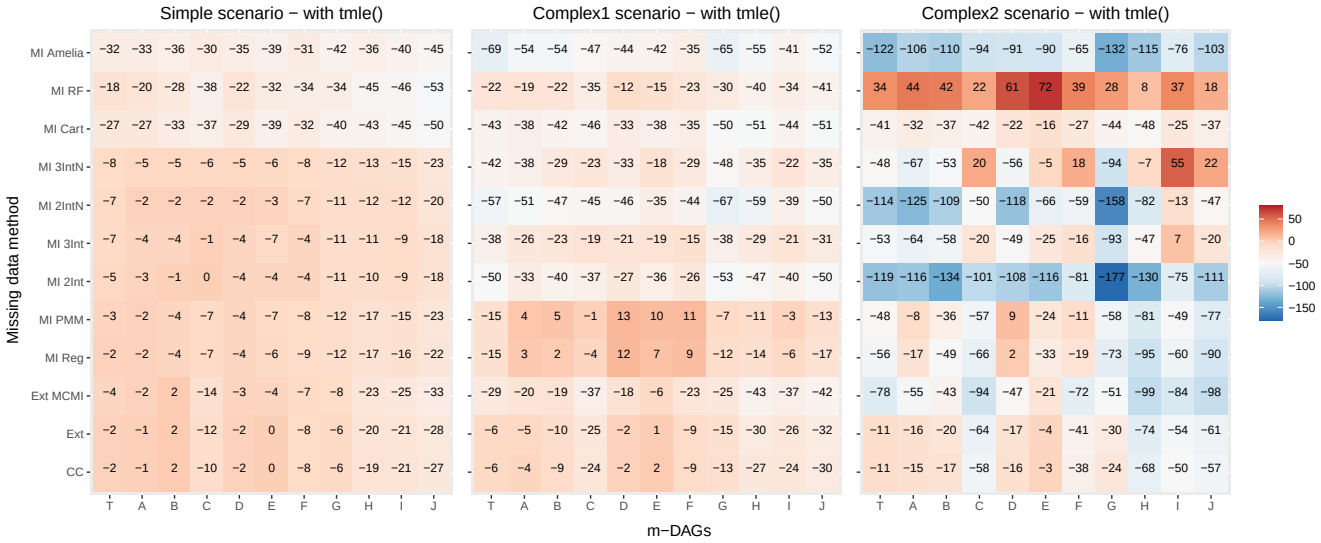


Figure 20: Relative bias (%) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE_S2* as described in chapter 4.5 in copula 2 data.

The relative bias is highest for the non-MI methods in the m-DAGs H, I, J. In the simple and complex 1 scenarios, the bias is low for the CC and Ext methods in the recoverable m-DAGs (figure 20). Ext MCMI exhibits a significantly larger bias across all m-DAGs compared to CC and Ext in the complex 1 and complex 2 scenarios. Additionally, in comparison, MI PMM and MI Reg also exhibit low bias. Generally, the bias is low for the parametric MIs. No clear trend can be observed for all missing data methods in the complex 2 scenario. Across all scenarios, MI Amelia exhibits the largest bias.

5.5.3 Relative error in ModSe

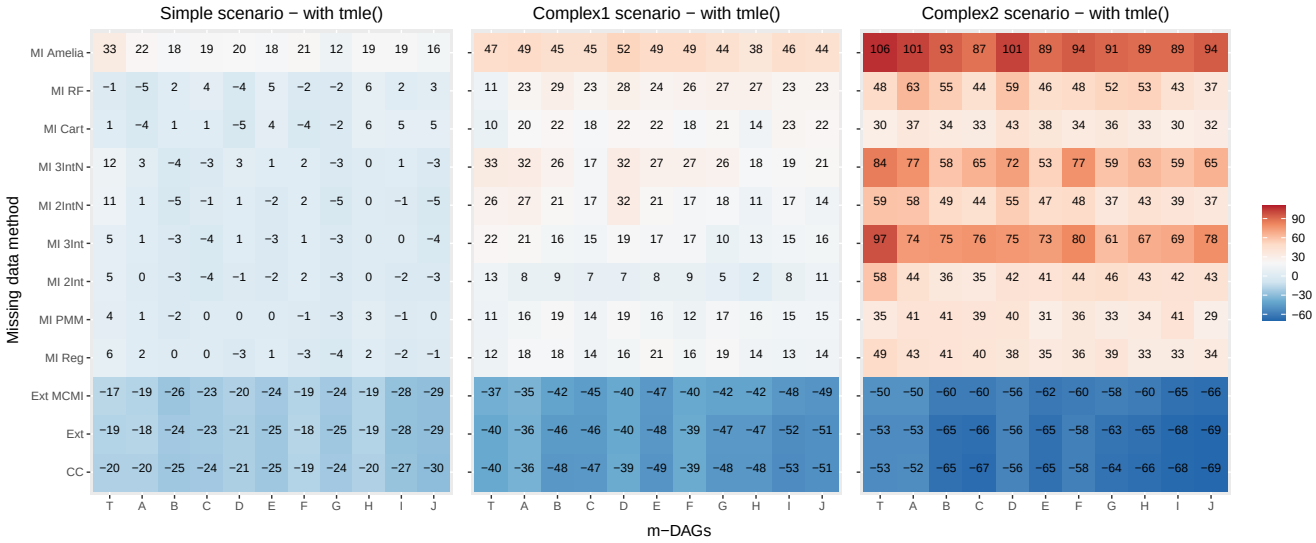


Figure 21: Relative error in ModSE (%) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE_S2* as described in chapter 4.5 in copula 2 data.

For simple and complex 1 scenarios, MI 2Int has the lowest error in ModSE among all m-DAGs (figure 20). MI Amelia overestimates ModSE the most compared to the other MI methods for all scenarios and m-DAGs. For simple and complex 1 scenarios, the error in ModSE is also evident for the non-MI methods. The underestimation of ModSE increases with the complexity of the scenarios for the non-MI methods. Furthermore the underestimation is very homogeneous across all m-DAGs.

5.5.4 RMSE

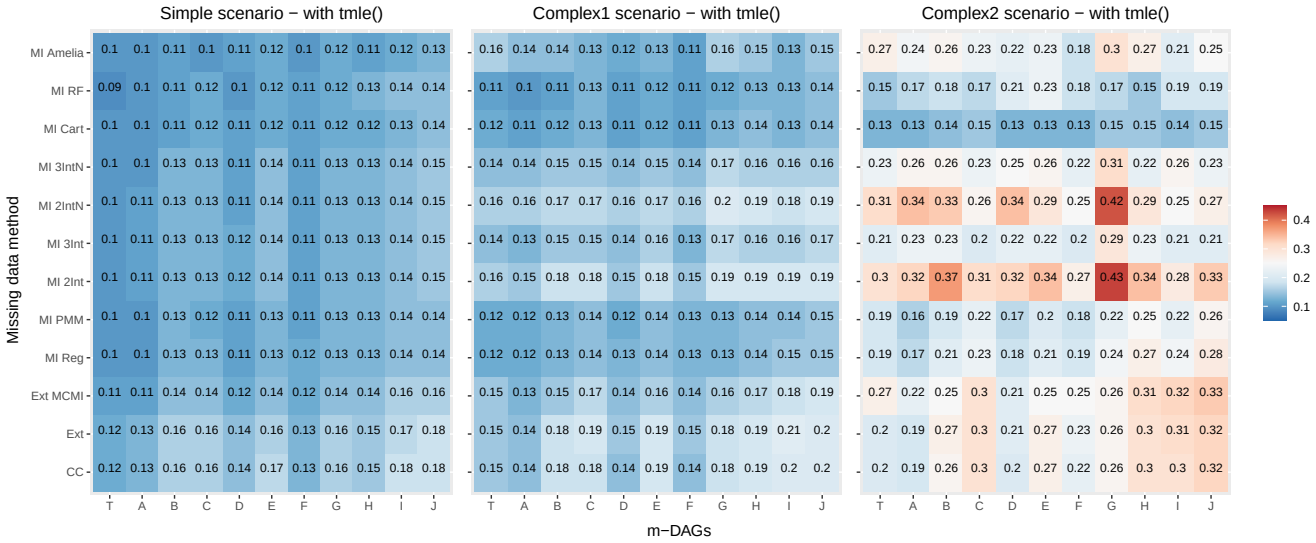


Figure 22: Root mean squared error (RMSE) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE_S2* as described in chapter 4.5 in copula 2 data.

For all scenarios MI CART has the lowest RMSE compared to the other missing methods (figure 22). In complex 2 scenario MI Amelia has high RMSE because of the high bias. In general, the non-MI methods have a higher RMSE across all scenarios and m-DAGs.

5.5.5 Coverage

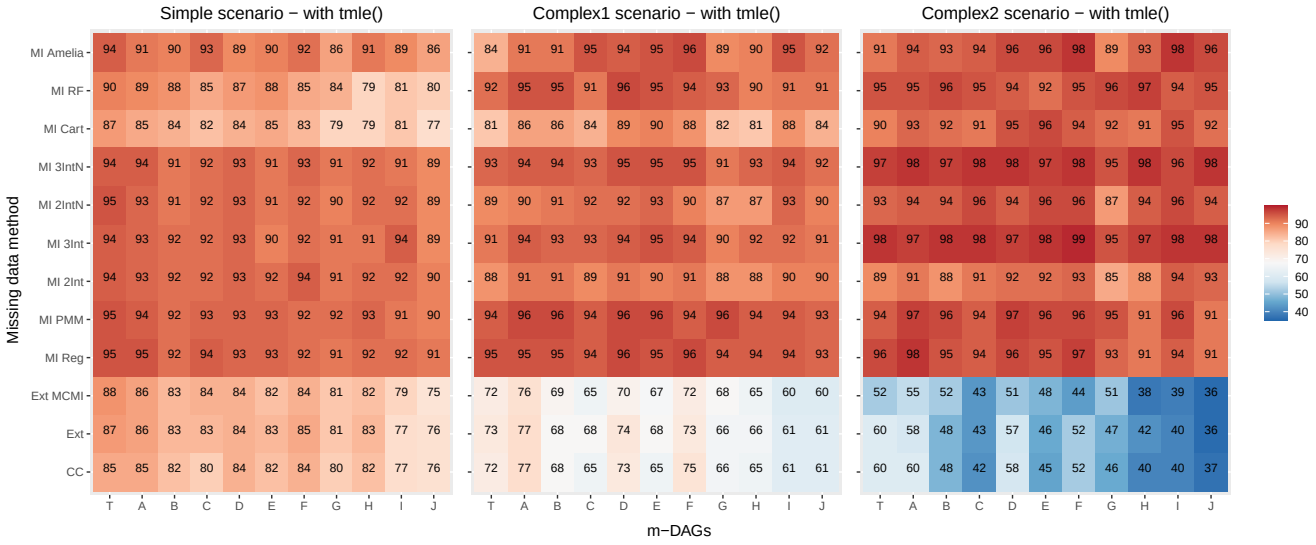


Figure 23: Coverage in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE_S2* as described in chapter 4.5 in copula 2 data.

The coverage probabilities of the 95% CI are displayed in figure 23 for the estimated ATE across all m-DAGs and considered scenarios in copula 2 data. For the non-MI methods, the coverage probabilities were reasonably acceptable in the simple scenario, except for the m-DAGs I, J. In the complex scenarios, the coverage deteriorated further and had poor coverage in the complex 2 scenario. In contrast, for the MI methods, the coverage was high again for all scenarios, which can be attributed to the overestimation of the ModSE.

5.6 Results of copula 3 data

5.6.1 Pre-analysis

	Estimated mean	Bias	Relative Bias (%)	empSE	RMSE	ModSE	Error in ModSE (%)	Coverage	Bias eliminated coverage	Mean CI length	Proportional CI length
Simple scenario											
TMLE	0.20	0.00	0.33	0.09	0.09	0.08	-14.86	0.88	0.88	0.30	0.92
TMLE_S2	0.20	-0.00	-0.25	0.09	0.09	0.08	-16.72	0.88	0.88	0.29	0.89
CVTMLE3	0.20	0.00	0.73	0.09	0.09	0.08	-12.29	0.91	0.90	0.32	1.00
TMLE3	0.20	0.00	1.15	0.10	0.10	0.07	-24.78	0.83	0.83	0.28	0.86
Complex 1 scenario											
TMLE	0.18	-0.02	-10.77	0.11	0.11	0.08	-26.03	0.82	0.82	0.31	0.87
TMLE_S2	0.17	-0.03	-14.96	0.10	0.11	0.07	-28.84	0.79	0.82	0.28	0.79
CVTMLE3	0.22	0.02	10.09	0.13	0.13	0.09	-28.89	0.83	0.84	0.36	1.00
TMLE3	0.26	0.06	30.35	0.15	0.16	0.07	-54.06	0.61	0.63	0.26	0.72
Complex 2 scenario											
TMLE	0.17	-0.03	-16.26	0.17	0.17	0.09	-43.36	0.68	0.71	0.35	0.91
TMLE_S2	0.16	-0.04	-19.76	0.16	0.16	0.09	-45.21	0.67	0.70	0.32	0.84
CVTMLE3	0.27	0.07	34.95	0.22	0.23	0.10	-53.74	0.60	0.63	0.39	1.00
TMLE3	0.36	0.16	79.60	0.25	0.29	0.07	-70.78	0.40	0.42	0.28	0.73

Table 9: Results for various TMLE implementations on complete copula 3 data across all scenarios

Table 9 represents the results of the performance of different TMLE implementations across all scenarios for the copula 3 data. The results for the simple scenario are similar to those obtained from the Copula 2 data for all performance measures, despite the error in ModSE which is roughly by 5 percentage points more underestimated for all TMLE implementations. As the scenario becomes more complex, the relative bias increases, with *TMLE3* consistently exhibiting the highest bias. Interestingly, the implementation of the **tmle3** package tends to overestimate the ATE, whereas *TMLE* and *TMLE_S2* underestimate the ATE. Additionally, the empirical standard error empSE increases for all TMLE implementations as the complexity of the scenarios increases. However, the poor coverage is observed for *CVTMLE3* and *TMLE3*.

5.6.2 Bias

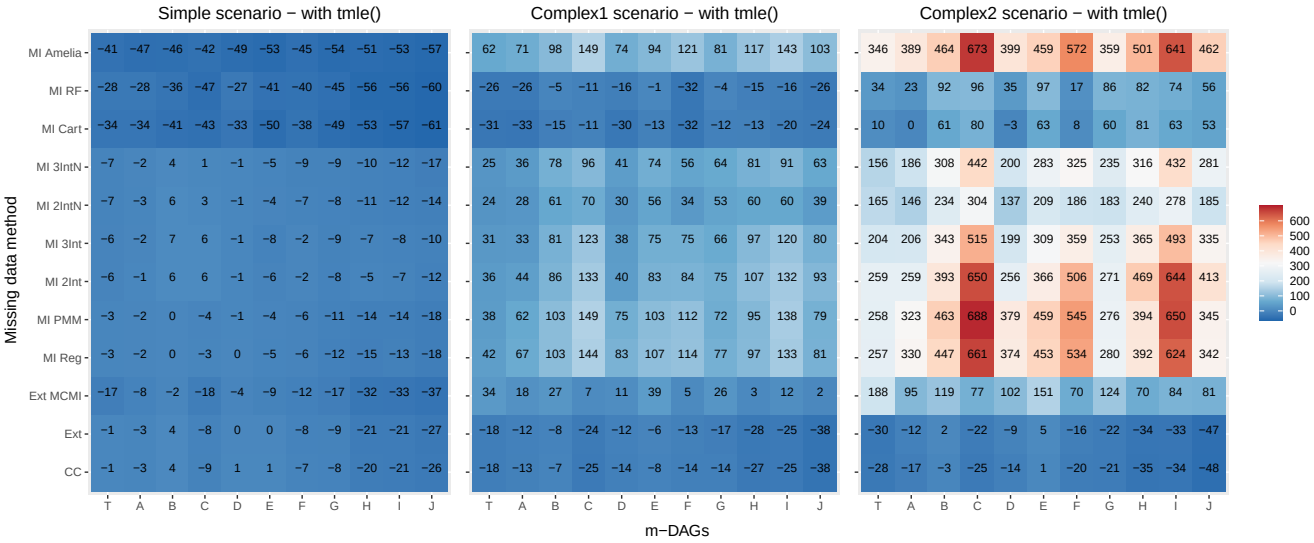


Figure 24: Relative bias (%) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE_S2* as described in chapter 4.5 in copula 3 data.

In figure 24, the relative bias of the estimated ATE of the presented missing data methods in combination with *TMLE* is shown.

For CC and Ext, the bias for m-DAGs H, I, J was again the largest in all scenarios. The bias for m-DAG C was also larger in the more complex scenarios. For complex 1 and complex 2 sceanrio, the bias for CC and Ext was the smallest across all m-DAGs. Ext MCMI exhibited a significantly larger bias across all m-DAGs compared to CC and Ext in the complex 2 scenarios. In the simple scenario, the bias for MI Amelia, MI RF, and MI CART was large for all m-DAGs, being the largest compared to the other missing methods.

5.6.3 Relative error in ModSe

For both simple and complex 1 scenarios, and across all m-DAGs, the error in ModSE was relatively small for all MI methods except MI Amelia (see Figure 25). MI Amelia consistently showed the highest overestimation of ModSE compared to the other MI methods in all scenarios and m-DAGs, except for the complex 2 scenario. In addition, for the simple and complex 1 scenarios, the error in ModSE was also observed for the non-MI methods. The underestimation of ModSE becomes more pronounced as the scenarios become more complex for the non-MI methods. It should be noted that all MI methods tend to overestimate ModSE in the complex scenarios.

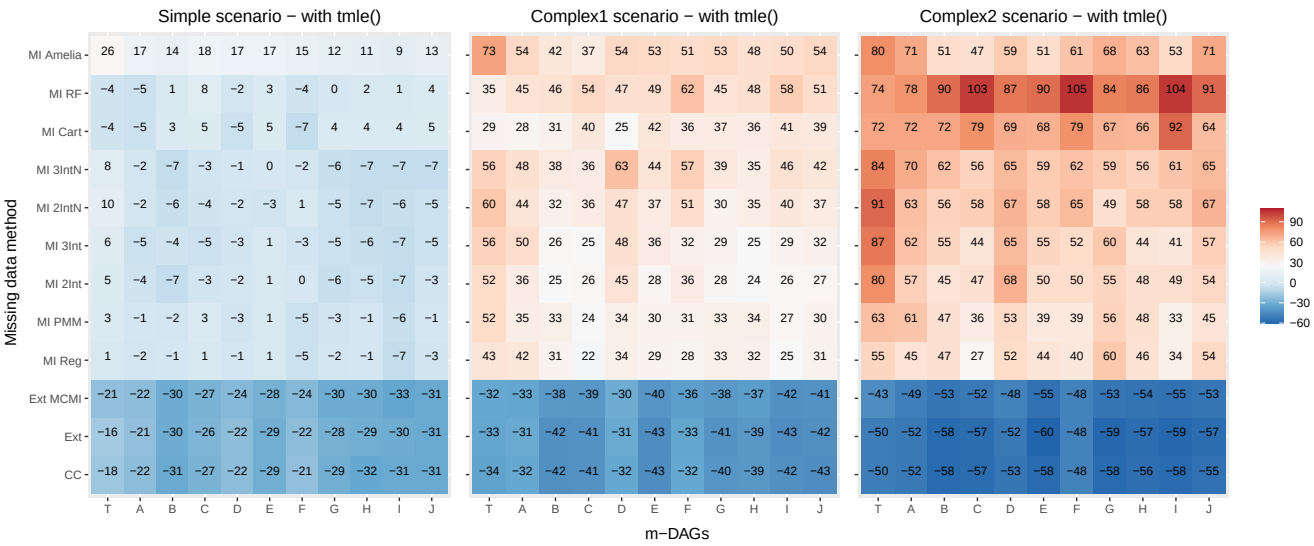


Figure 25: Relative error in ModSE (%) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE_S2* as described in chapter 4.5 in copula 3 data.

5.6.4 RMSE

For both complex 1 and complex 2 scenarios, CC and Ext exhibit the lowest root mean square error (RMSE) among the other missing data imputation methods (see figure 26), except for MI CART and MI RF. The parametric MI methods and MI Amelia display a particularly high RMSE for complex 2 scenario due to their significant bias. However, in the simple scenario, the MI methods outperform the non-MI methods.

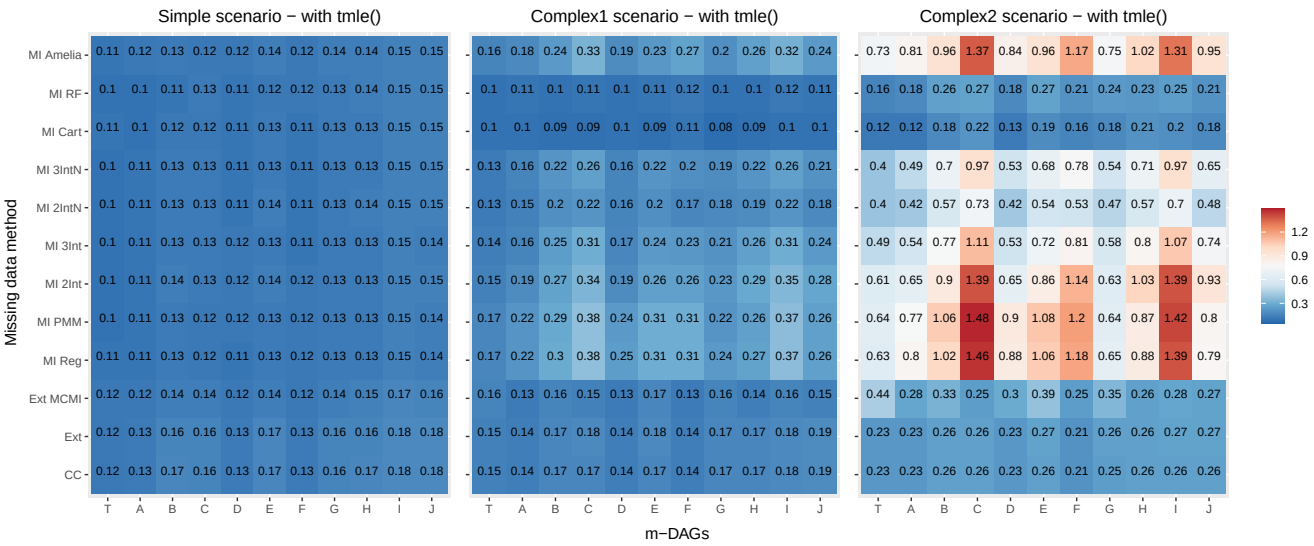


Figure 26: Root mean squared error (RMSE) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE_S2* as described in chapter 4.5 in copula 3 data.

5.6.5 Coverage

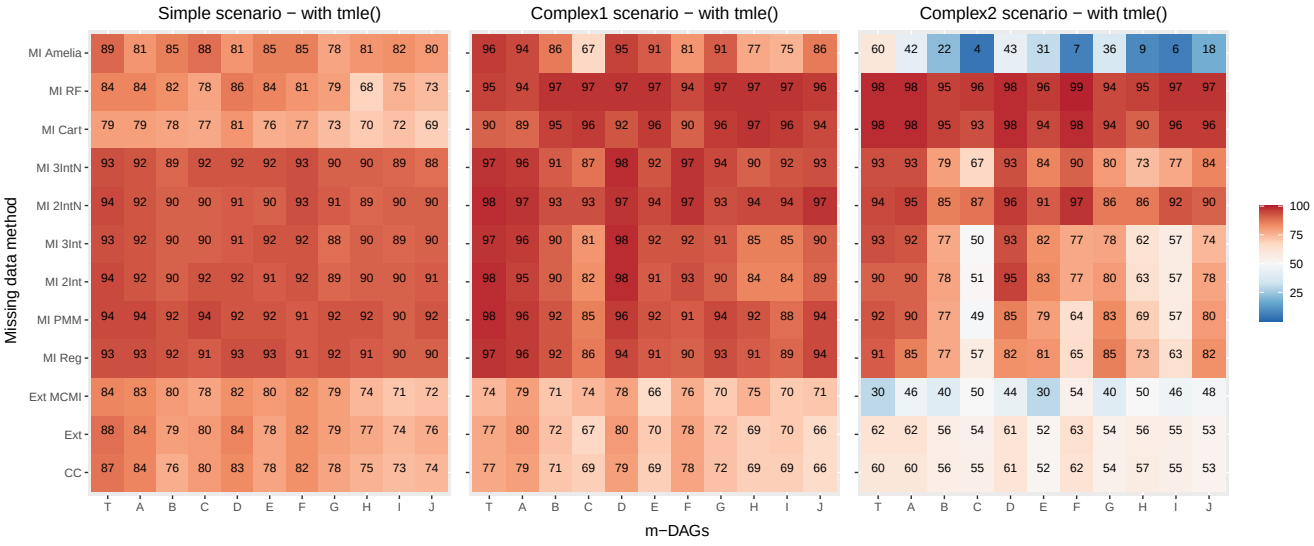


Figure 27: Coverage in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE_S2* as described in chapter 4.5 in copula 3 data.

Due to the significant underestimation of the ModSe, once again, a lower coverage is observed for the CC and Ext methods. In complex 2 scenario, MI Amelia had the lowest coverage (see figure 27).

6 Discussion

The findings of the reproduced data indicate that the effectiveness of various missing data imputation methods can differ significantly depending on the particular scenario and m-DAG being examined. No single method was found to consistently outperform all others across all m-DAGs in terms of relative bias. Nevertheless, relatively low bias was observed for m-DAGs T, A, B, D, E, F, G when the non-MI methods, CC and Ext were utilized. For more complex m-DAGs H, I, and J, the MI method using Amelia demonstrated lower relative bias in the complex 2 scenario. In scenarios where missingness did not depend on the outcome for any variable and where a complete-case analysis could recover the conditional distribution of the outcome (m-DAGs T, A, B, D, E), low bias was observed when using the CC and Ext methods. This was noted regardless of the complexity of the data generation procedure. The same methods maintained relatively low bias even for m-DAGs F and G, where the conditional distribution of the outcome isn't recoverable. A slightly higher bias compared to m-Dag F was observed for m-DAG G, which had missingness in the outcome directly influenced by the outcome. Ext MCMI method, however, resulted in more biased estimates across these scenarios. When applied to the other non-recoverable m-DAGs H, I, J, an slight decrease in bias compared to other non-MI methods CC and Ext was observed.

In the simple scenario, the parametric MI methods with no interactions, MI PMM and MI Reg, was found to generate smaller bias when the outcome did not directly influence missingness in the outcome (m-DAGs T, A, B, C, D, E, F, I) and higher bias when the outcome did directly influence missingness in the outcome (G, H, J). In the complex scenarios, a trend towards higher bias was observed with these MI methods with no interactions. Especially for m-DAGs T, A, B, C, D, E, F bias was high in complex 2 scenario.

The inclusion of interaction terms in the imputation models, namely MI 2Int, MI 3Int, MI 2IntN and MI 3IntN appeared to improve the performance of parametric MI methods in these complex scenarios. The findings suggest that MI 2Int, MI 3Int, MI 2IntN and MI 3IntN, which incorporate interaction terms, performed well in complex scenarios because these scenarios were defined based on the presence and intensity of confounder-confounder interactions. This means that MI models with interactions were reasonably well-specified to handle the data structure of these scenarios. However, it's crucial to remember that in real-world observational studies, the underlying process that generates the data is often not fully understood. This ambiguity makes it difficult to define parametric imputation models that are perfectly compatible with the models used for analysis. Additionally, for datasets with significantly more variables, it is difficult and tedious to include all interaction effects in **mice** package. It's also important to consider that some interactions could be collinear and in general, the estimates could be unstable (van Buuren and Groothuis-Oudshoorn, 2011). This difficulty highlights the value of using a data-adaptive approach such as MI CART, also because MI CART led to significantly lower bias compared MI RF. Another convenient application is offered by MI Amelia, demonstrating a lower bias for the non-recoverable m-DAGs F, G, H, I, J, especially in more complex scenarios. Despite this advantage, it's worth noting that MI CART and MI Amelia did not uniformly reduce bias across all types of m-DAGs and all three scenarios, and were not superior compared to the non-MI methods like

CC and Ext.

Regarding the results of the reproduced data, it is seen that for non-MI methods the ModSEs are underestimated, and this underestimation increases with the complexity of the scenarios. This might be due to the fact that these methods ignore the uncertainty associated with missing data, leading to overly optimistic estimates of the model error. This was also the case for all different TMLE implementations performed using the complete data and was thus not surprising. On the other hand, MI methods tend to overestimate the ModSE, and this overestimation increases as the complexity of the scenario increases. This could be because MI methods take into account the uncertainty associated with the imputation process, which can result in larger estimates of the model error. Particularly, methods that include two-way or additionally three-way and four-way interactions (MI 2Int, MI 3Int, MI 2IntN, and MI 3 IntN) and more flexible methods like CART and Amelia seem to be more sensitive to increasing complexity, likely because they can capture more complex patterns in the data that might not actually be useful or accurate for the imputation. Additionally, the performance of the MI Rubin variance estimator is anticipated to be suboptimal in situations of incompatibility, as suggested in prior research (Bartlett and Hughes, 2020). This may provide an explanation for the observed error ModSE associated with the MI approaches in the present study. It is suggested that incompatibility presents a significant challenge when utilizing MI in conjunction with TMLE, where models are fitted using Super Learner.

It can be observed that for MI using Amelia, the empSE and RMSE are the lowest across all m-dags and scenarios. Although MI Amelia tends to overestimate the ModSE to some extent, this overestimation is not problematic when considering the length of the CI and comparing it to other MI methods or even to the non-MI methods, CC and Ext, which underestimate the ModSE and still have a similar CI length. Nevertheless, it should be mentioned that overestimation of ModSE leads to overcoverage (Morris et al., 2019). In practical terms, this implies that the MI Amelia method has performed well in this analysis, demonstrating both precision and accuracy in handling missing data across a variety of scenarios and data structures represented by different m-DAGs. Its ability to estimate with low empSE and high coverage make it a reliable choice for dealing with missing data in this context. The superior performance of MI Amelia could potentially be attributed to the nature of the DGP. The reproduced data only contains binary confounders and exposure. With binary variables, the uncertainty introduced by missingness is often less than that introduced by continuous variables, as there are only two possible values for imputation. Therefore, if the missing values are conditional on the categories of the binary variables, it might be easier to assign the missing values accurately to their respective classes. Amelia´s EMB algorithm makes fewer assumptions about the distribution of the data, which can provide more robust results against model mis-specification. In other words, even if the actual data distribution does not align with the distribution assumed by the model, Amelia can often still perform well (Honaker et al., 2011).

The results for the positivity violation data broadly confirm the findings from the reproduced data. However, the bias for the non-MI methods is now higher for the non-recoverable m-DAGs F and G. Furthermore, it can be noted that the RMSE for the non-MI methods CC and Ext is significantly

worse compared to the MI methods. This is not due to a larger bias, but rather to a higher dispersion (empSE). For MI Amelia, the bias was now significantly higher for all m-Dags and scenarios, but still had a lowest RMSE overall due to the high precision (low empSE). Furthermore the inclusion of interaction terms in the imputation models, namely MI 2Int, MI 3Int, MI 2IntN and MI 3IntN no longer appeared to improve the performance of parametric MI methods in both complex scenarios under high positivity violation (40%).

For the modified, copula 1, copula 2, and copula 3 data, the bias was generally low for the recoverable m-DAGs T, A, B, D, E, and slightly increased for non-recoverable m-DAGs F, G for the methods CC and Ext. The method Ext MCMI consistently had a higher bias and was particularly distorted in the complex 2 scenario. Regarding the complex 2 scenario, it can be noted that all MI methods had a significantly higher bias compared to CC and Ext. Interestingly, even the MI methods MI 2Int, MI 3Int, MI 2IntN, and MI 3IntN displayed a relatively large bias in the complex 2 scenarios for these expanded datasets. These MI methods should theoretically be well-specified and capable of accurately capturing the relationships in the data. However, an underperformance compared to CART was apparent. One main reason could be that the interactions in the data are extremely complex, and even correctly specified interaction terms might not fully encapsulate the complexity of these relationships. Tree-based methods like CART can sometimes better capture these complex interaction structures as they can create more intricate partitions of the data. Furthermore, compared to the reproduced data, there are now continuous confounders in the dataset, thereby making outliers possible. Linear regression, which some of these MI methods may use, can be sensitive to outliers, which might unduly influence the imputed values.

Generally, the MI methods exhibited higher distortion in the complex 1 and complex 2 scenarios but had much lower empSE and consequently lower RMSE compared to the non-MI methods CC and Ext. A possible reason for this is that, MI methods produce in general lower empSEs because they use the correlations between variables to generate more accurate and consistent estimates of the missing values. This reduces the spread or dispersion of the estimates across the multiple imputed datasets, resulting in a lower empSE. However, these methods rely on certain assumptions, and if these assumptions are violated, they may produce biased results (van Buuren, 2018). On the other hand, in non-MI methods, cases with any missing values are completely dropped, which can result in loss of power or increase in empSE.

The good performance of MI Amelia on the reproduced data and also on the positivity violation data could not be confirmed for the expanded datasets.

Overall, it can be concluded that CC and Ext methods have relatively low bias except for the m-DAGs C, H, I, J. However, they possess a larger empSE than the MI methods and additionally underestimate the ModSe, which in turn leads to poor coverage. When considering the RMSE, for the more complex scenarios and extended datasets, MI CART has performed the best.

For the different TMLE implementations, some surprising results were obtained. First of all, across all considered datasets except positivity violation data, and across all scenarios, the implementation from the package **tmle** yielded the lowest bias, the lowest empSE, and therefore the lowest RMSE. It is also

evident that while ModSE is consistently underestimated across all scenarios, its underestimation is still lower compared to the other implementations from the **tmle3** and **ltmle** packages, resulting in not only higher coverage but also relatively narrower CIs.

It is clear that *TMLE* and *TMLE_S2* provided the best performance compared to the other TMLE implementations for the given simulations. Additionally, in all different data scenarios, it was not evident that *TMLE*, which utilizes the *cv.Qinit* argument and is designed to protect against overfitting by cross-validating the initial Super Learner estimate of $\bar{Q}_0(A, W)$, provided better estimates than *TMLE_S2*. Both implementations produced identical results, making *TMLE_S2* more attractive for application due to its approximately halved runtime, as the additional layer of cross-validation is eliminated.

The implementations *LTMLE* and *LTMLE_ic* showed a significant bias regarding the estimation of ATE for reproduced data, positivity violation data, and modified data. This bias logically increased with the complexity of the scenarios. Compared to the other TMLE implementations, the empSE was similarly large and the underestimation of ModSE was also similar to *TMLE* and *TMLE_S2*. The high bias is probably more due to the Super Learner implementation than to the inclusion of the clever covariates as weights in the logistic regression for updating the initial estimates. Especially when outcome Y is continuous and needs to be rescaled accordingly, the results from the **ltmle** package deviate from those of the **tmle**. Surprisingly, there are no performance differences between *LTMLE* and *LTMLE_ic* in terms of estimating ModSE. Particularly in the default setting, the robust variance, an approach that directly targets the variance of the influence function as a counterfactual mean outcome (Tran et al., 2018), and the influence curve-based variance are estimated, and the larger of the two is used. Furthermore, the estimation using **ltmle** takes by far the longest, and the runtime increases even further when continuous covariates are used instead of just binary variables. The required runtime makes it nearly impossible to conduct larger simulation studies with **ltmle**.

Lastly, for the **tmle3** package, the *TMLE3* implementation, especially in more complex scenarios across different datasets, exhibited a high underestimation of the ModSE in comparison to other implementations. Moreover, the empSE was relatively large, resulting in a combination of low coverage while maintaining relatively large confidence intervals (CI). The bias for this was significantly high only in the complex scenarios for copula 1 and copula 3 data. Nonetheless, *CVTMLE3* provided a lower bias across all scenarios and datasets. For positivity violation data, the distortion was even less than for *TMLE* and *TMLE_S2*. Compared to *TMLE3*, the underestimation of ModSE was almost always less. For implementation using the **tmle3**, this at least confirms the theory that, for valid statistical inference, cross-validation (CV) is beneficial (Balzer and Westling, 2021). This is mainly because the initial estimator of $\bar{Q}_0(A, W)$ is prone to overfitting. Thus, there is no realistic residual variation left for the targeting step, making the update incapable of mitigating residual bias (van der Laan and Zheng, 2010), (Gruber and van der Laan, 2012). However, both the underestimation of ModSE and the bias, except in positivity violation data, are greater than for *TMLE* and *TMLE_S2*. Therefore, it would, in principle, be possible and necessary to try to manually combine CV with *TMLE_S2* to more accurately verify the benefits of CV-TMLE. One advantage of *CVTMLE3* over *TMLE3* is its marginally longer

runtime, despite the fact that it essentially uses an additional layer of cross-validation.
A weakness in the used execution of missing data methods in combination with TMLE in this simulation study is that the missing data methods are not applied once, but rather for each TMLE implementation separately. This introduces additional uncertainty in estimating the ATE. However, it can be observed through the implementations *TMLE* and *TMLE_S2*, both of which have nearly identical performance, that the variability remained within limits. Nevertheless, it is advisable to address first missingness in the datasets and then apply the different TMLE implementations to estimate the ATE.
Another interesting point is, for example, using the different TMLE implementations, it is possible to compare and determine the biases regarding the ATE on the full dataset. Now, considering the distance between the biases of the individual TMLE implementations, it is observed that this distance relatively decreases in relation to the different TMLE implementations in combination with MI methods, while it remains unchanged for the non-MI methods.

7 Conclusion

In conclusion, for the proposed missing data methods and the implementation *TMLE*, the results of the simulation study (Dashti et al., 2021) were consistently replicated.
The analysis of reproduced data demonstrates the variability in performance across different missing data imputation methods, contingent on the specific scenario and m-DAG under consideration. There was no single method that consistently excelled across all m-DAGs in terms of relative bias. However, lower relative bias was detected for certain m-DAGs T, A, B, D, E, F, G when non-MI methods, specifically CC and Ext, were used. For more intricate m-DAGs H, I, J, MI Amelia performed admirably with lower relative bias in complex scenarios.
The CC and Ext method performed well where missingness did not depend on the outcome for any variable and a complete-case analysis could recover the conditional distribution of the outcome. These methods maintained relative low bias for non-recoverable m-DAGs F and G, even though a slightly higher bias was observed for m-DAG G, which exhibited missingness in the outcome directly influenced by the outcome. In contrast, Ext MCMI method exhibited more biased estimates across these scenarios.
The results indicate that the non-MI methods underestimated the ModSEs, and this underestimation increased with the complexity of the scenarios. This could be because these methods ignore the uncertainty associated with missing data, leading to overly optimistic estimates of the model error. In contrast, MI methods tended to overestimate the ModSE, and this overestimation increases as the complexity of the scenario increases.
The MI Amelia method has shown to perform well in this analysis, demonstrating both precision and low bias (leading to low RMSE) in handling missing data across a variety of scenarios and data structures represented by different m-DAGs. The superior performance of MI Amelia could potentially be attributed to the nature of the data generation process.
Results from the positivity violation data largely confirm the outcomes from the original reproduced

data, with a few differences worth noting. The bias for non-MI methods was now more pronounced for the non-recoverable m-DAGs F and G. Additionally, MI Amelia exhibited an increased bias across all m-DAGs and scenarios, despite maintaining the lowest overall RMSE due to high precision.
In terms of the modified, copula 1, copula 2, and copula 3 data, the bias was generally low for the recoverable m-DAGs and slightly higher for non-recoverable m-DAGs using methods CC and Ext. Moreover, for these expanded datasets, all MI methods demonstrated a significantly higher bias in the complex 2 scenario compared to CC and Ext. This suggests that even correctly specified interaction terms might not fully encapsulate the complexity of these relationships, and methods like CART may be more adept at capturing these complex interaction structures.
While MI methods presented higher bias in the complex scenarios, they exhibited lower empirical standard errors (empSE) and thus lower RMSE compared to non-MI methods.
Notably, the superior performance of MI Amelia on the reproduced data and positivity violation data was not confirmed for the expanded datasets.
In summary, the CC and Ext methods tend to demonstrate relatively low bias, except for m-DAGs C, H, I, J. However, these methods have higher empSE than the MI methods and tend to underestimate the ModSe, resulting in poor coverage. Considering the RMSE, MI CART outperforms other methods in more complex scenarios and extended datasets.
In terms of the different TMLE implementations, it was found that the TMLE implementations *TMLE* and *TMLE_S2* package yielded the lowest bias, the lowest empSE, and therefore the lowest RMSE across all scenarios. On top of that *TMLE_S2* has by far lowest runtime, making this approach superior. Meanwhile, implementations such as *LTMLE* and *LTMLE_ic* exhibited a significant bias for estimating ATE.
Finally, the *TMLE3* implementation, especially in more complex scenarios across different datasets, exhibited a high underestimation of ModSE in comparison to other implementations. However, *CVTMLE3* provided a lower bias and underestimation of ModSE across all scenarios and datasets in comparison to *TMLE3*, could possibly confirm the theory that, for valid statistical inference CV is beneficial and could add robustness.
Therefore, this comprehensive analysis underscores the importance of considering the characteristics of the specific scenario and m-DAG when choosing an imputation method, the intricacies of different implementations of TMLE, and the impact of increasing complexity on the reliability of imputation and estimation methods.

List of Figures

List of Tables

List of Abbreviations

MCAR	Missing Completely at Random
MAR	Missing at Random
MNAR	Missing Not at Random
DAG	Directed Acyclic Graph
c-DAG	complete Directed Acyclic Graph
m-DAG	missing-data Directed Acyclic Graph
MI	Multiple Imputation
JM	Joint Modeling
FCS	Fully Conditional Specification
MICE	Multiple Imputation by Chained Equations
MLE	Maximum Likelihood Estimation
EMB	Expectation-Maximization with Bootstrapping
EM	Expectation-Maximization
NNLS	Non-Negative Least Squares
TMLE	Targeted Maximum Likelihood Estimation
ATE	Average Treatment Effect
SUTVA	Stable Unit Treatment Value Assumption
NPSEM	Non-Parametric Structural Equation Models
DGD	Data Generating Distribution
CV-TMLE	Cross-Validated Targeted Maximum Likelihood Estimation
IC	Influence Curve
EmpSE	Empirical Standard Error
RMSE	Root Mean Squared Error
ModSE	Model Standard Error
CI	Confidence Interval
DGP	Data Generating Process
VAHCS	Victorian Adolescent Health Cohort Study
CC	Complete Case Analysis
Ext	Extended Targeted Maximum Likelihood Estimation
Ext MCMI	Extended TMLE plus missing covariate missing indicator approach
CIT	Conditionally Independent Treatment
CIO	Conditionally Independent Outcomes
SE	Standard Error
MI REG	Parametric Multiple Imputation using linear regression
PMM	Predictive Mean Matching
MI PMM	Parametric Multiple Imputation using Predictive Mean Matching
MI 2Int	Parametric Multiple Imputation with two-way Interactions
MI 3Int	Parametric Multiple Imputation with additionally three-way and higher Interactions

MI 2IntN	Parametric Multiple Imputation with defaults and two-way Interactions
MI 3IntN	Parametric Multiple Imputation with defaults and additionally three-way and higher Interactions
MI CART	Multiple Imputation using Classification and Regression Trees
CART	Classification and Regression Trees
MI RF	Multiple Imputation using Random Forest
CV	Cross Validation
Pos1	Positvity violation 1 (30%)
Pos2	Positvity violation 2 (40%)

References

Balzer, L. B. and Westling, T. (2021). Demystifying statistical inference when using machine learning in causal research. *American Journal of Epidemiology*.

Bartlett, J. and Hughes, R. (2020). Bootstrap inference for multiple imputation under uncongeniality and misspecification. *Statistical Methods in Medical Research*.

Blake, H. A., Leyrat, C., Mansfield, K. E., Tomlinson, L. A., Carpenter, J., and Williamson, E. J. (2020). Estimating treatment effects with partially observed covariates using outcome regression with missing indicators. *National Library of Medicine*.

Coyle, J. (2017). Computational considerations for targeted learning. *UC Berkeley*.

Coyle, J. (2023). *tmle3: The Extensible TMLE Framework*. R package version 0.2.0.

Coyle, J., Hejazi, N., Malenica, I., Phillips, R., and Sofrygin, O. (2023). *sl3: Pipelines for Machine Learning and Super Learning*. R package version 1.4.4.

Dashti, S. G., Lee, K. J., Simpson, J. A., White, I. R., Carlin, J. B., and Moreno-Betancur, M. (2021). Handling missing data when estimating causal effects with targeted maximum likelihood estimation. *arXiv*, 1.

Diaz, I. and van der Laan, M. J. (2017). Doubly robust inference for targeted minimum loss-based estimation in randomized trials with missing outcome data. *National Library of Medicine*.

Doove, L., Van Buuren, S., and Dusseldorp, E. (2014). Recursive partitioning for missing data imputation in the presence of interaction effects. *Computational Statistics and Data Analysis*.

Greenland S., Pearl J., R. J. (1999). Causal diagrams for epidemiologic research. *National Library of Medicine*.

Gruber, S. and van der Laan, M. (2012). tmle: An R package for targeted maximum likelihood estimation. *Journal of Statistical Software*, 51(13):1–35. doi:10.18637/jss.v051.i13.

Hernán, M. A. and Robins, J. M. (2020). *Causal Inference: What If*. Chapman Hall/CRC.

Hofert, M., Kojadinovic, I., Maechler, M., and Yan, J. (2023). *copula: Multivariate Dependence with Copulas*. R package version 1.1-2.

Holland, P. W. (1986). The causal foundations of structural equation modeling. *Journal of the American Statistical Association*.

Honaker, J., King, G., and Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, 45(7):1–47.

Karthika, M. and Judea, P. (2019). Graphical models for processing missing data. *arXiv*.

Kurz, C. F. (2021). Augmented inverse probability weighting and the double robustness property. *National Library of Medicine*.

Lendle, S. D., Schwab, J., Petersen, M. L., and van der Laan, M. (2017). ltmle: An R package implementing targeted minimum loss-based estimation for longitudinal data. *Journal of Statistical Software*, 81(1):1–21.

Li, H., Rosete, S., Coyle, J., Phillips, R. V., Hejazi, N. S., Malenica, I., Arnold, B. F., Benjamin-Chung, J., Mertens, A., Colford, J. M. J., van der Laan, M. J., and Hubbard, A. E. (2022). Evaluating the robustness of targeted maximum likelihood estimators via realistic simulations in nutrition intervention trials. *National Library of Medicine*.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. Wiley Online Library.

Luque-Fernandez, M. A., Schomaker, M., Rachet, B., and Schnitzer, M. (2018). Targeted maximum likelihood estimation for a binary treatment: A tutorial. *Statistics in medicine*.

Montoya, L., Skeem, J., van der Laan, M., and Petersen, M. (2021). Performance and application of estimators for the value of an optimal dynamic treatment rule. *arXiv*.

Moreno-Betancur, M., Lee, K. J., Leacy, F. P., White, I. R., Simpson, J. A., and Carlin, J. B. (2018). Canonical causal diagrams to guide the treatment of missing data in epidemiologic studies. *American Journal of Epidemiology*, 187.

Morris, T., White, I., and Crowther, M. (2019). Using simulation studies to evaluate statistical methods. *National Library of Medicine*.

Morris, T., White, I. R., and Royston, P. (2014). Tuning multiple imputation by predictive mean matching and local residual draws. *National Library of Medicine*.

Naimi, A. I. and Balzer, L. B. (2018). Stacked generalization: An introduction to super learning. *National Library of Medicine*.

Nelsen, R. B. (2006). An introduction to copulas. *Springer*.

Pearl, J. (2010). The causal foundations of structural equation modeling. *Technical Report R-370*.

Petersen, M. L., Porter, K. E., Gruber, S., Wang, Y., and van der Laan, M. J. (2012). Diagnosing and responding to violations in the positivity assumption. *National Library of Medicine*.

Polley, E., LeDell, E., Kennedy, C., and van der Laan, M. (2021). *SuperLearner: Super Learner Prediction*. R package version 2.0-28.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rubin, D. B. (1975). Inference and missing data. *Wiley Online Library*.

Rubin, D. B. (1977). Formalizing subjective notion about the effect of nonrespondents in sample surveys. *Journal of the American Statistical Association*, 72.

Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. *Wiley Online Library*, 72.

Rubin, D. B. (2005). Causal inference using potential outcomes. *Journal of the American Statistical Association*.

Schuler, M. S. and Rose, S. (2016). Targeted maximum likelihood estimation for causal inference in observational studies. *National Library of Medicine*.

Tran, L., Petersen, M., Schwab, J., and van der Laan, M. J. (2018). Robust variance estimation and inference for causal effect estimation. *arXiv*.

van Buuren, S. (2018). *Flexible Imputation of Missing Data*, volume 2. Chapman & Hall/CRC.

van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3):1–67.

van der Laan, M. J., Polley, E., and Hubbard, A. (2007). Super learner. *Statistical applications in genetics and molecular biology*.

van der Laan, M. J. and Rose, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer.

van der Laan, M. J. and Rubin, D. (2006). Targeted maximum likelihood learning. *The International Journal of Biostatistics*.

van der Laan, M. J. and Zheng, W. (2010). Asymptotic theory for cross-validated targeted maximum likelihood estimation. *U.C. Berkeley Division of Biostatistics Working Paper Series*.

White, I. R. and Carlin, J. B. (2010). Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *National Library of Medicine*.

White, I. R., Royston, P., and Wood, A. M. (2010). Multiple imputation using chained equations: Issues and guidance for practice. *Wiley Online Library*.

A Figures

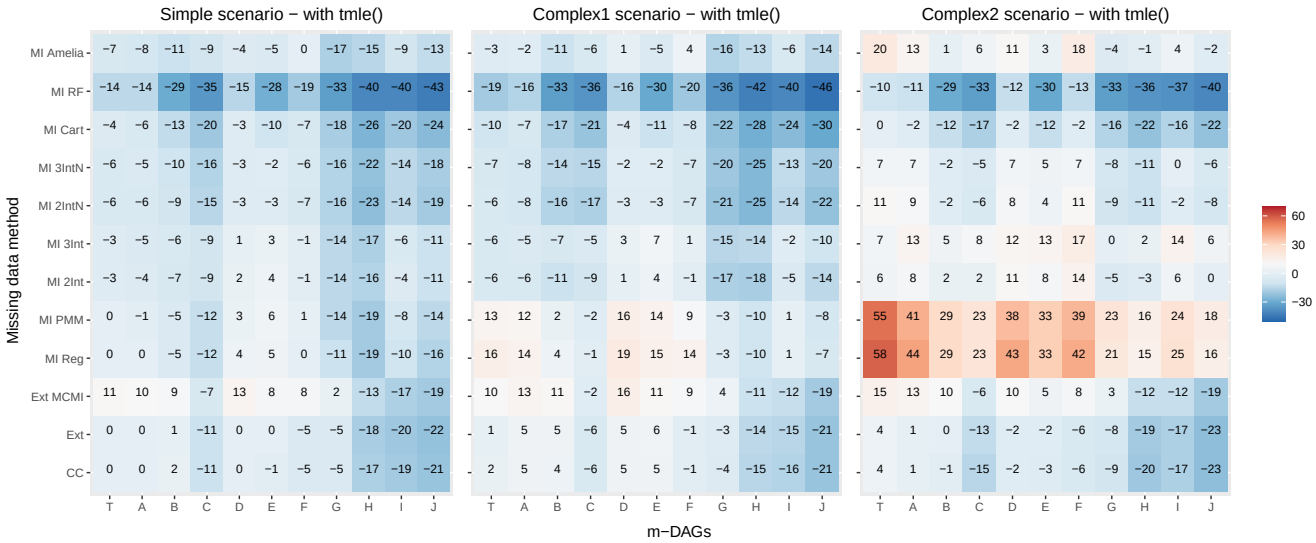


Figure A.1: Relative bias (%) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE_S2* as described in chapter 4.5 in reproduced data.

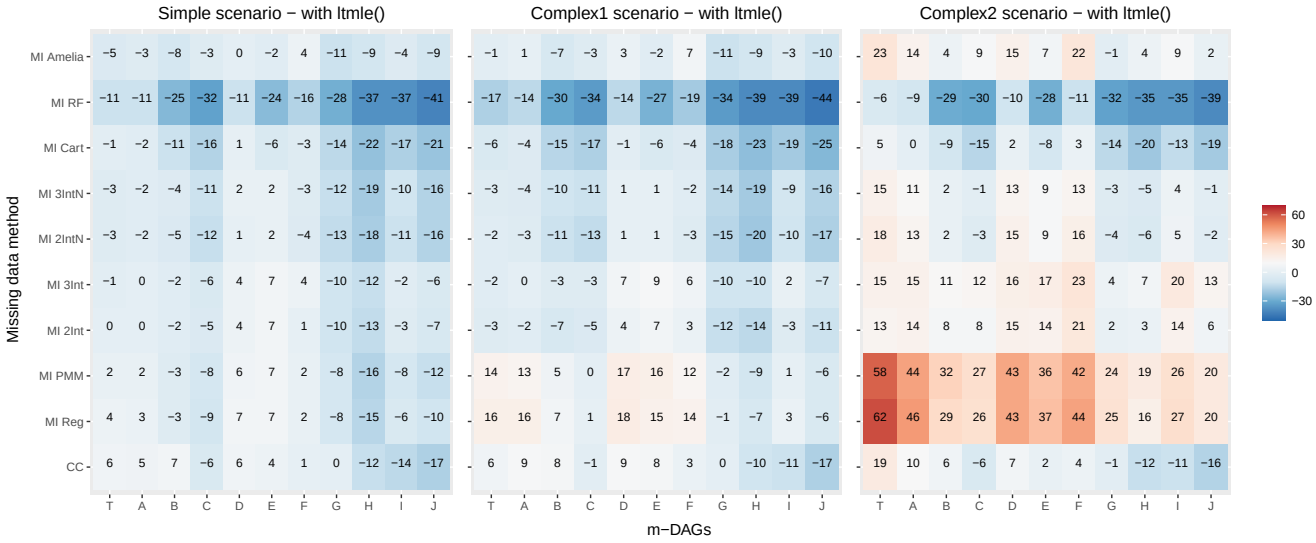


Figure A.2: Relative bias (%) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *LTMLE* as described in chapter 4.5 in reproduced data.

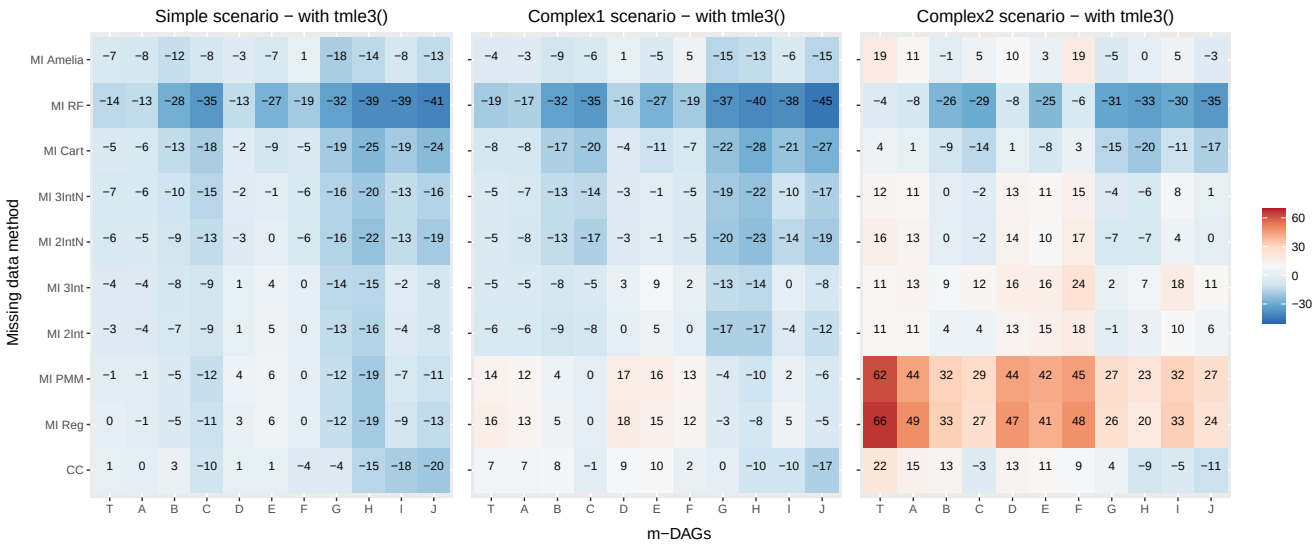


Figure A.3: Relative bias (%) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE3* as described in chapter 4.5 in reproduced data.

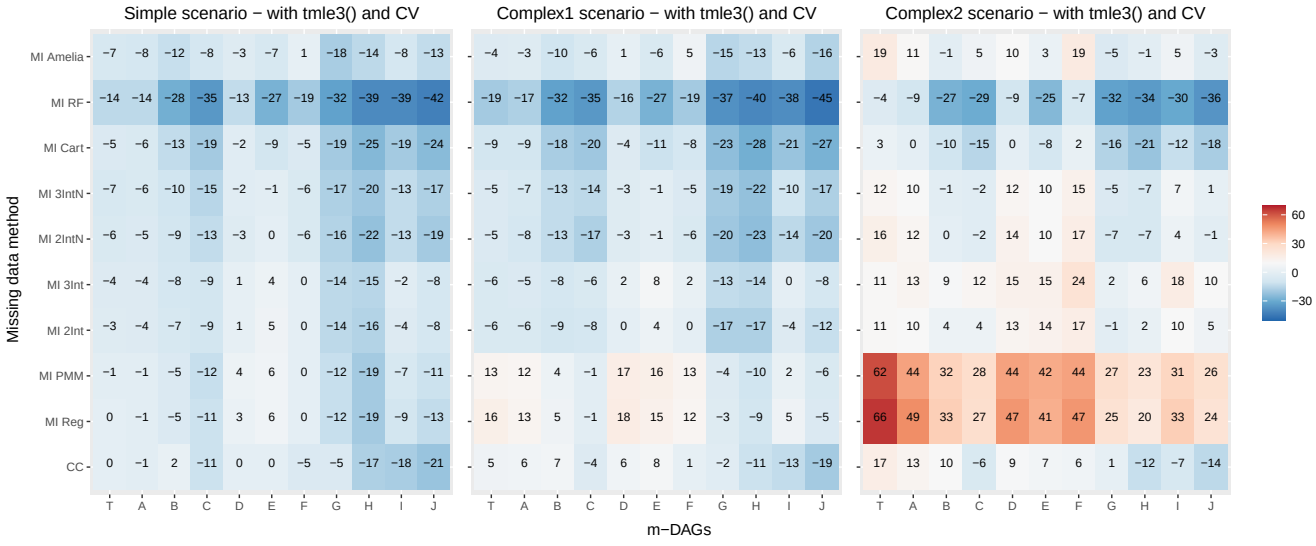


Figure A.4: Relative bias (%) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *CVTMLE3* as described in chapter 4.5 in reproduced data.

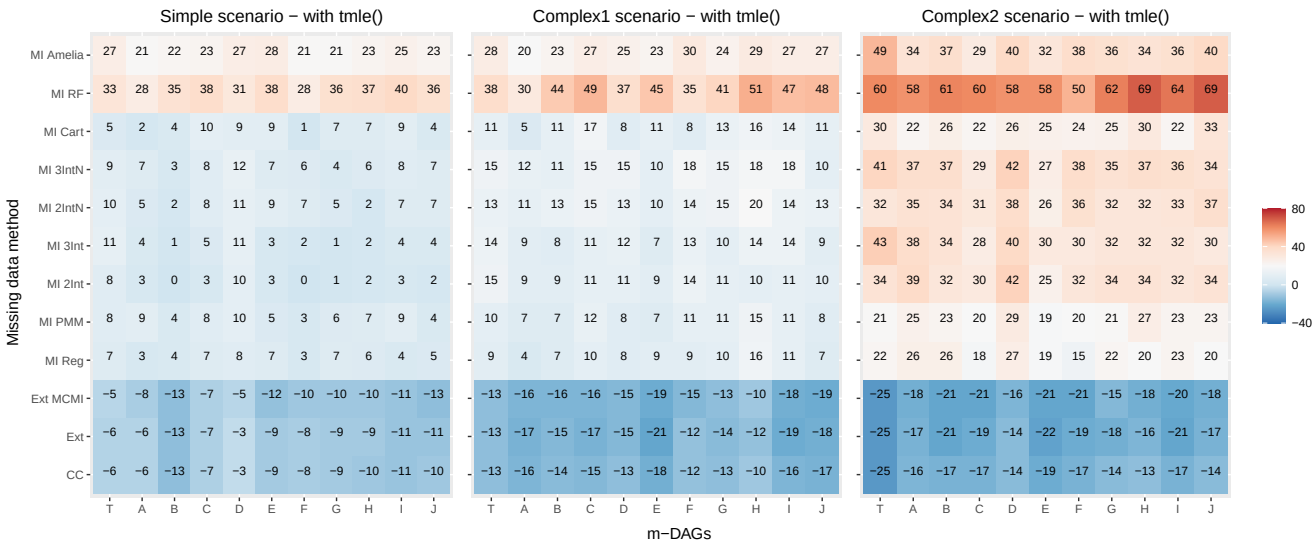


Figure A.5: Relative error in ModSE (%) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE_S2* as described in chapter 4.5 in reproduced data.

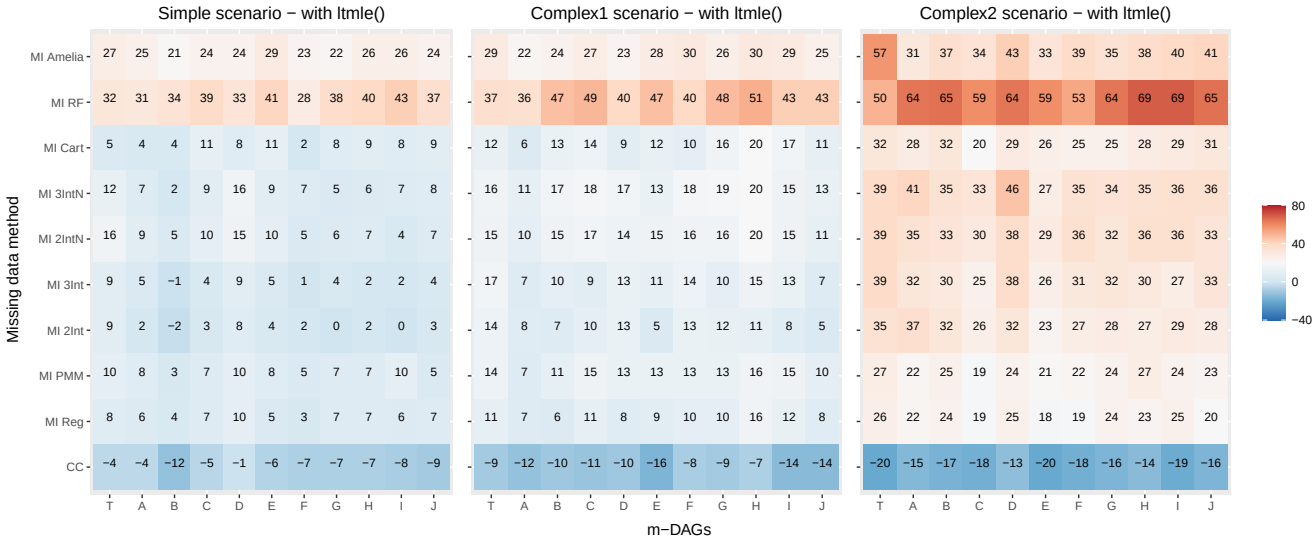


Figure A.6: Relative error in ModSE (%) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *LTMLE* as described in chapter 4.5 in reproduced data.

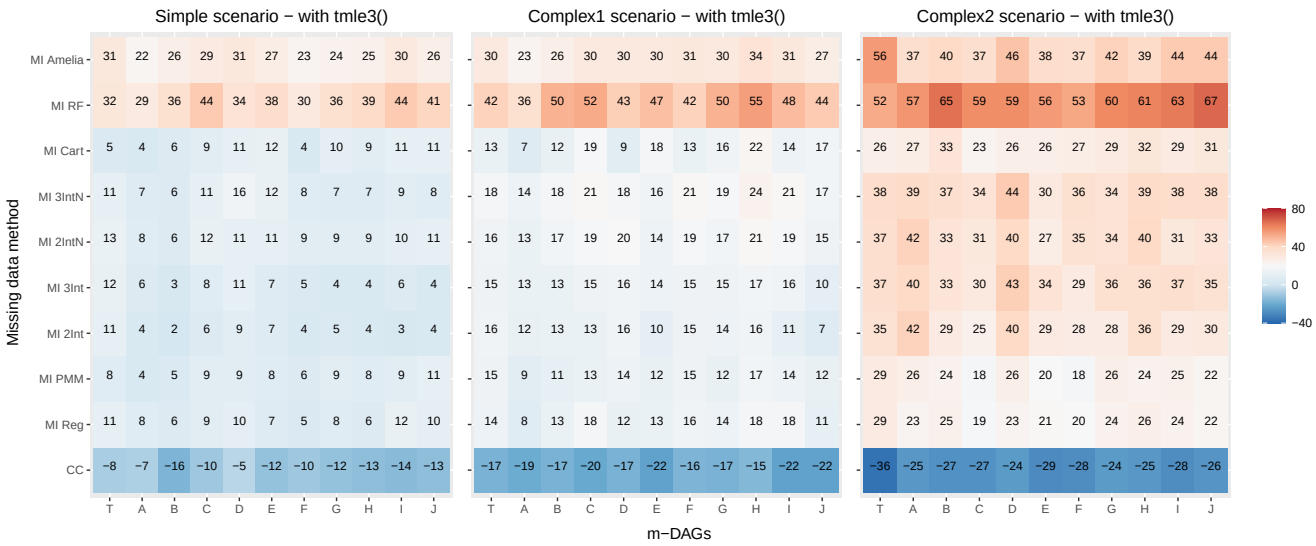


Figure A.7: Relative error in ModSE (%) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE3* as described in chapter 4.5 in reproduced data.

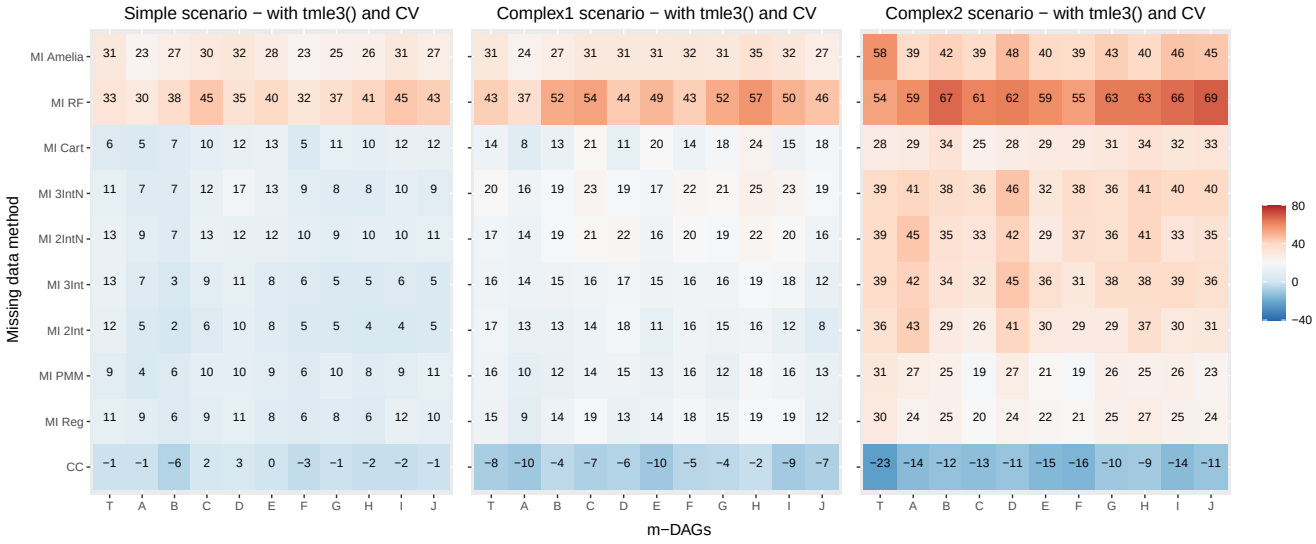


Figure A.8: Relative error in ModSE (%) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *CVTMLE3* as described in chapter 4.5 in reproduced data.

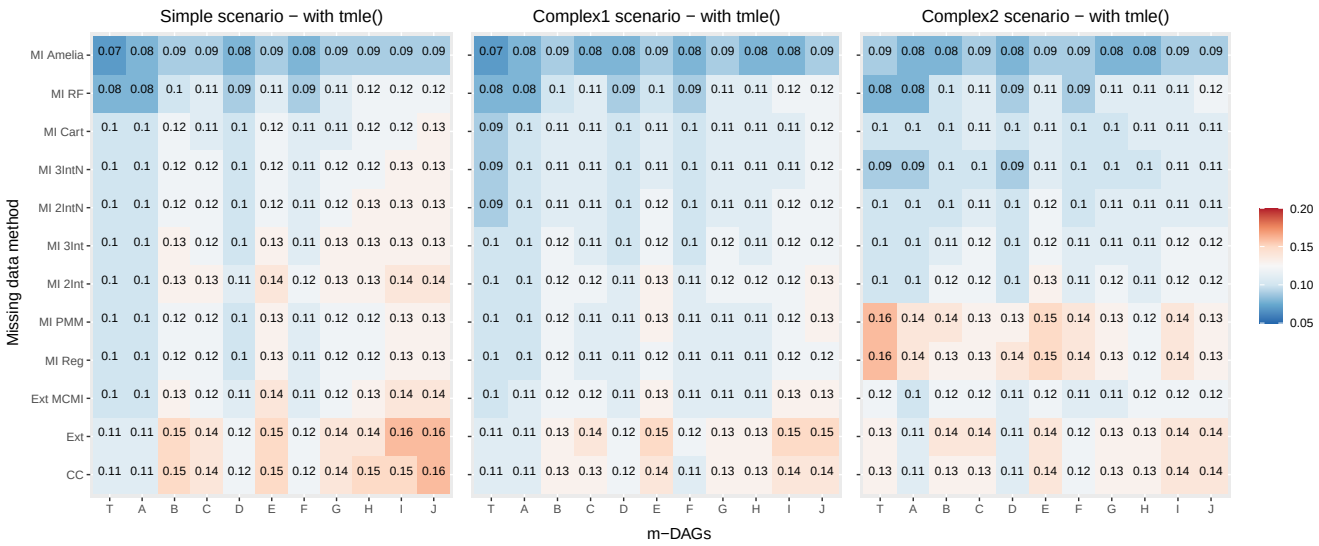


Figure A.9: Root mean squared error (RMSE) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE_S2* as described in chapter 4.5 in reproduced data.

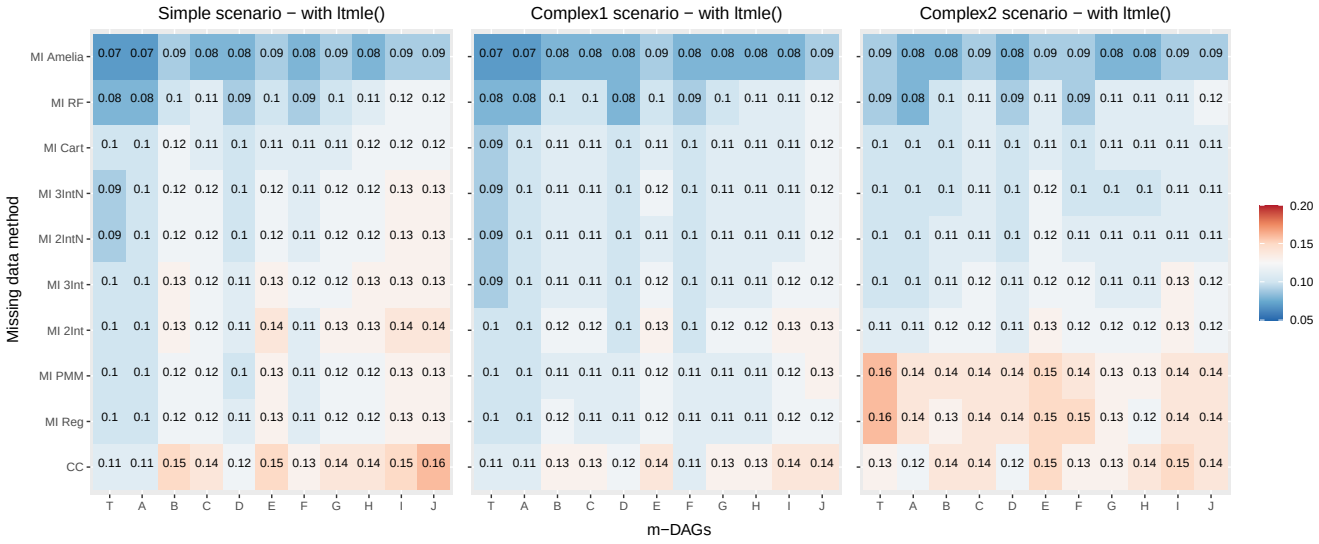


Figure A.10: Root mean squared error (RMSE) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *LTMLE* as described in chapter 4.5 in reproduced data.

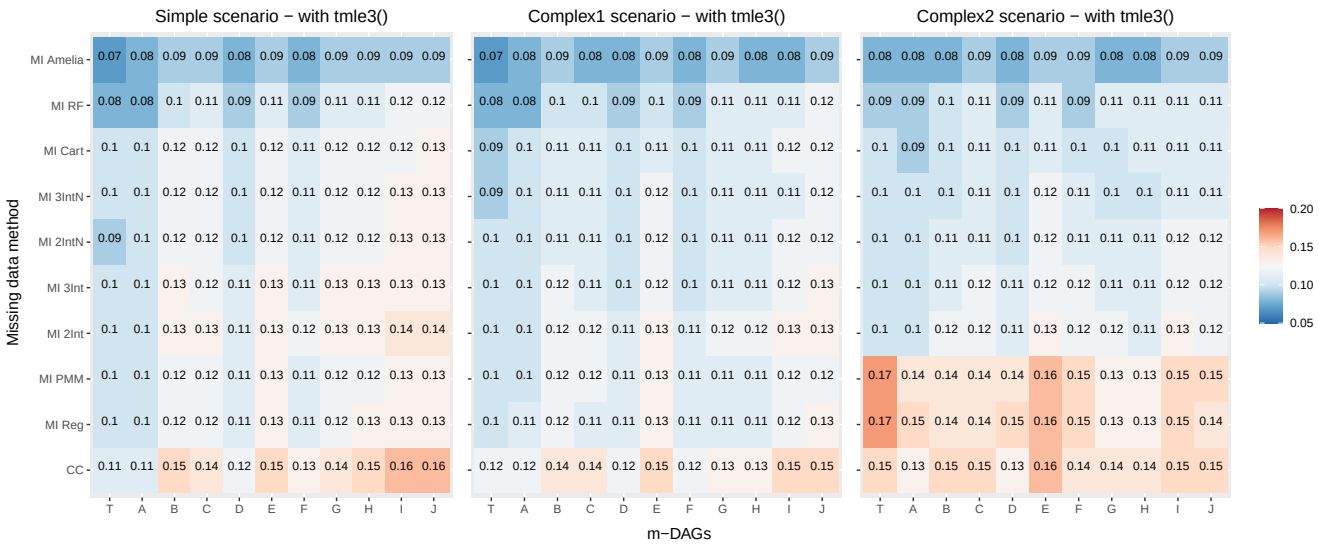


Figure A.11: Root mean squared error (RMSE) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE3* as described in chapter 4.5 in reproduced data.

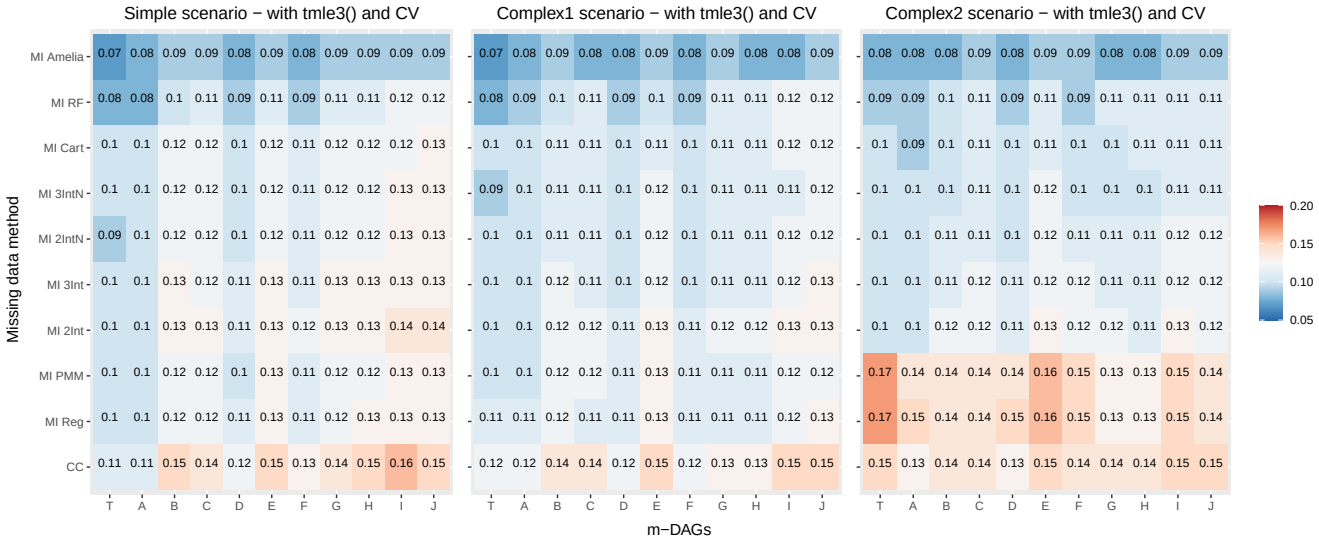


Figure A.12: Root mean squared error (RMSE) in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *CVTMLE3* as described in chapter 4.5 in reproduced data.

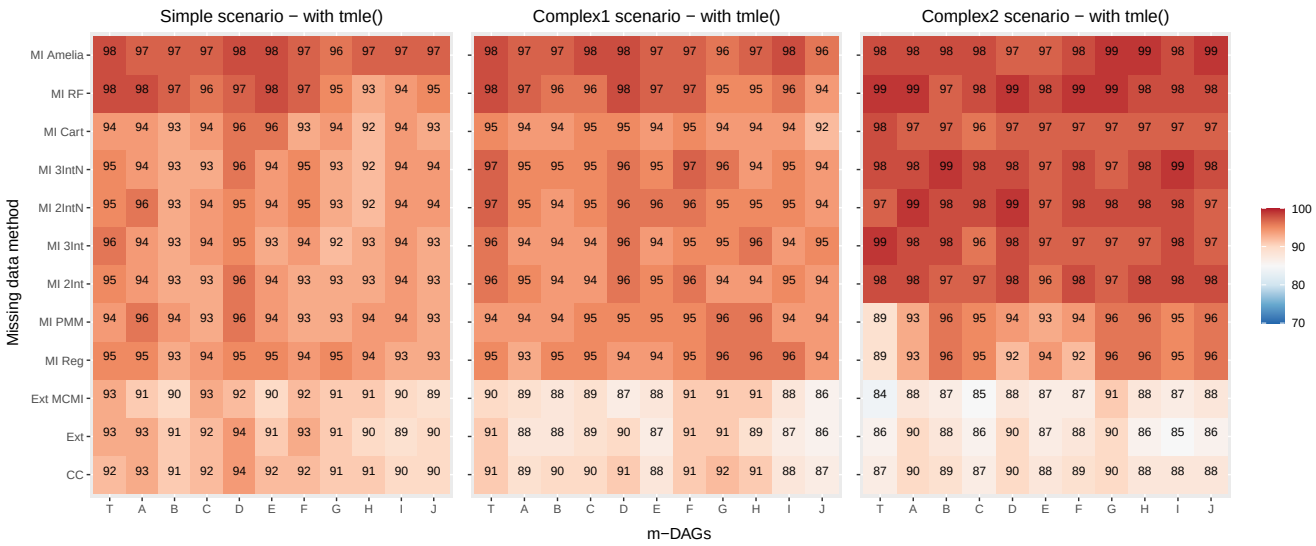


Figure A.13: Coverage in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE_S2* as described in chapter 4.5 in reproduced data.

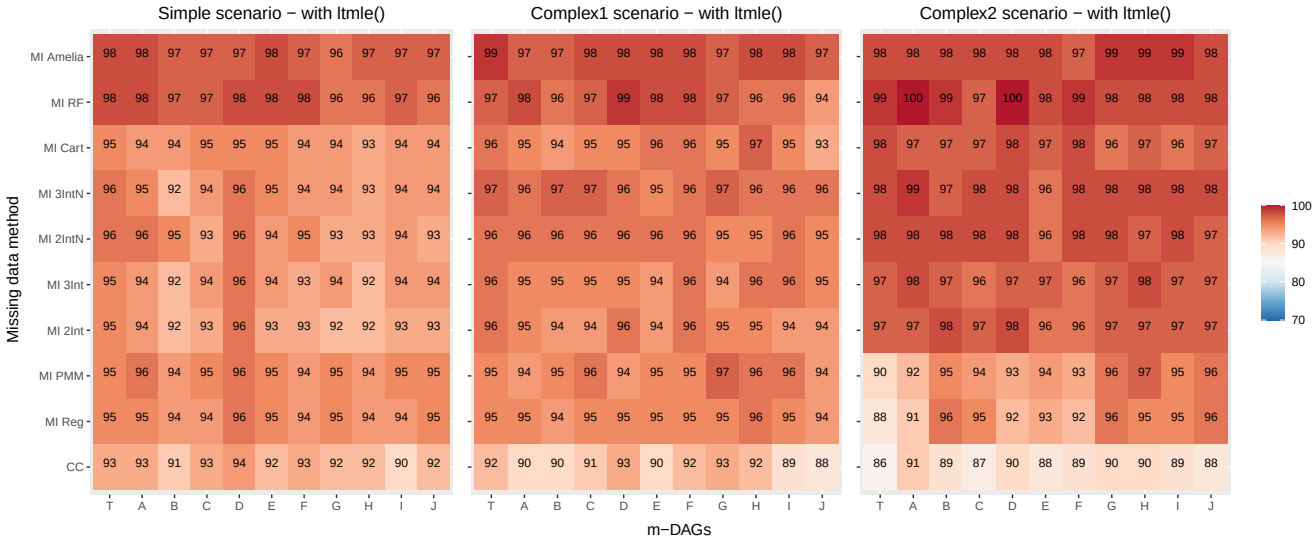


Figure A.14: Coverage in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *LTMLE* as described in chapter 4.5 in reproduced data.

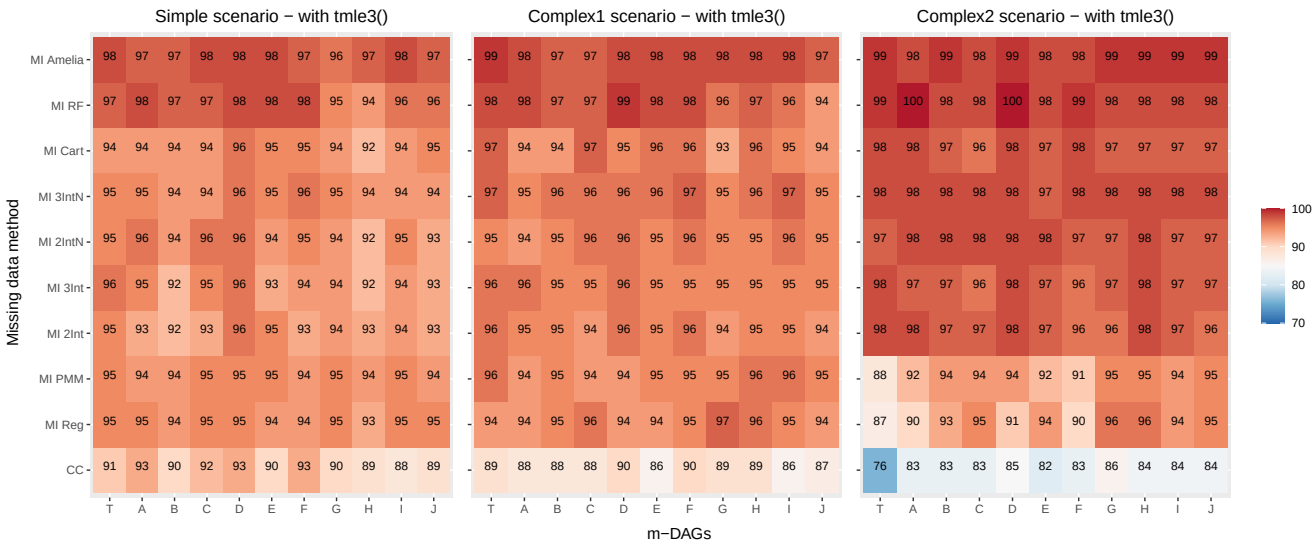


Figure A.15: Coverage in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *TMLE3* as described in chapter 4.5 in reproduced data.

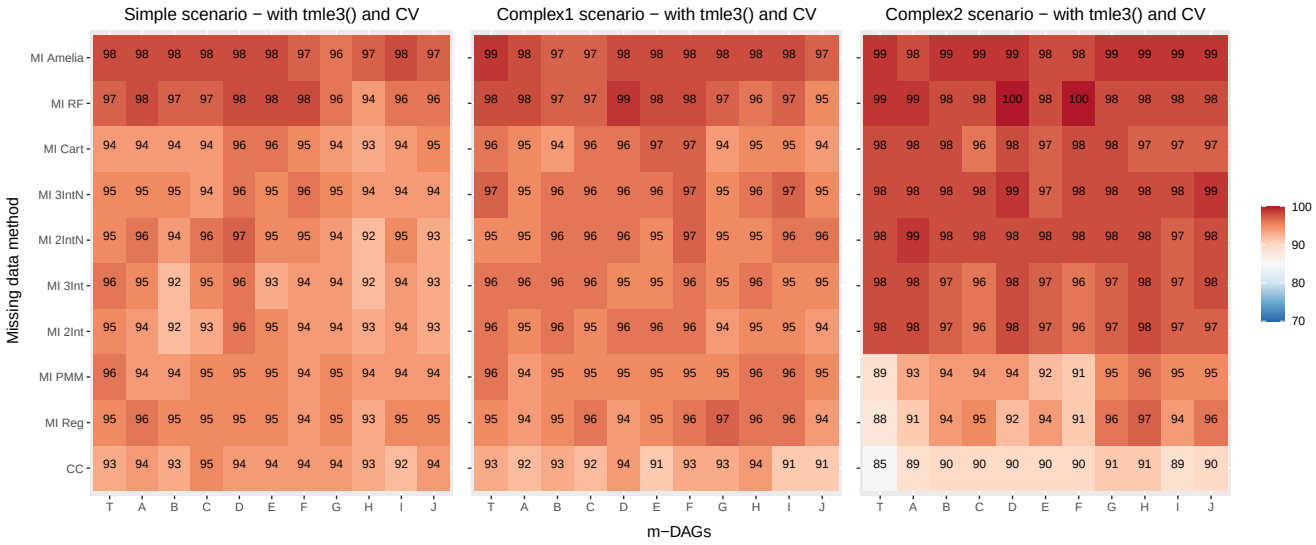


Figure A.16: Coverage in ATE estimation using different missing data methods for the 11 assessed missingness directed acyclic graphs (m-DAGs) and implementation *CVTMLE3* as described in chapter 4.5 in reproduced data.

B Tables

	Model for	Regression coefficient of												
		Intercept	W1	W2	W3	W4	W5	A	Y	B	MW2	MW3	MW4	MZA
Complete data	W1	-1.30												
	W2	-1.90								0.40				
	W3	0.40								0.70				
	W4	-0.60								-0.70				
	W5	-0.50												
	Both complexity and positivity violation scenarios													
	A		1.00	0.70	-2.00	-1.40	1.20			-2.40				
	Y		0.10	0.40	0.70	0.20	0.30	0.20						
DAG T	MW2	-0.85												
	MW3	-4.40									4.30			
	MW4	-4.00									3.90	1.40		
	MA	-2.00									1.50	1.50	1.50	
	MY	-1.60									-0.50	0.50	0.50	0.50
DAG A	MW2	-1.45	0.90				0.90							
	MW3	-5.60	0.90				0.90				4.80			
	MW4	-4.70	0.90				0.90				3.90	1.50		
	MA	-2.50	0.90				0.90				1.30	1.50	1.50	
	MY	-2.10	0.90				0.90				0.10	0.10	0.10	0.05
DAG B	MW2	-1.60	0.90				0.90	0.90						
	MW3	-5.20	0.90				0.90	0.90			4.10			
	MW4	-4.40	0.90				0.90	0.90			3.20	2.00		
	MA	-3.70	0.90	0.90	0.90	0.90	0.90				1.80	1.50	1.50	
	MY	-3.25	0.90	0.90	0.90	0.90	0.90	0.90			-0.30	0.10	0.10	0.10
DAG C	MW2	-1.60	0.90				0.90	0.90	0.10					
	MW3	-5.30	0.90				0.90	0.90	0.10		4.30			
	MW4	-4.50	0.90				0.90	0.90	0.10		3.50	1.30		
	MA	-3.70	0.90	0.90	0.90	0.90	0.90		0.10		1.70	1.50	1.50	
	MY	-3.25	0.90	0.90	0.90	0.90	0.90	0.90			-0.40	0.10	0.10	0.15
DAG D	MW2	-1.60	0.90	0.90			0.90							
	MW3	-6.20	0.90		0.90		0.90				4.80			
	MW4	-5.00	0.90			0.90	0.90				3.90	1.50		
	MA	-2.65	0.90				0.90	0.90			1.30	1.50	1.50	
	MY	-2.10	0.90				0.90				0.10	0.10	0.10	0.10
DAG E	MW2	-1.75	0.90	0.90			0.90	0.90						
	MW3	-5.70	0.90		0.90		0.90	0.90			4.10			
	MW4	-4.80	0.90			0.90	0.90	0.90			3.20	2.00		
	MA	-3.80	0.90	0.90	0.90	0.90	0.90	0.90			1.50	1.50	1.50	
	MY	-3.20	0.90	0.90	0.90	0.90	0.90	0.90			-0.60	0.10	0.10	0.20
DAG F	MW2	-1.60	0.90	0.90			0.90		0.10					
	MW3	-6.60	0.90		0.90		0.90		0.10		4.10			
	MW4	-5.40	0.90			0.90	0.90		0.10		3.20	2.00		
	MA	-2.55	0.90				0.90	0.90	0.10		1.20	1.30	1.30	
	MY	-2.10	0.90				0.90				-0.30	0.10	0.10	0.40
DAG G	MW2	-1.60	0.90				0.90	0.90						
	MW3	-5.20	0.90				0.90	0.90			4.10			
	MW4	-4.45	0.90				0.90	0.90			3.20	2.00		
	MA	-3.70	0.90	0.90	0.90	0.90	0.90				1.70	1.50	1.50	
	MY	-3.30	0.90	0.90	0.90	0.90	0.90	0.90	0.10		-0.30	0.10	0.10	0.10
DAG H	MW2	-1.60	0.90				0.90	0.90	0.10					
	MW3	-5.40	0.90				0.90	0.90	0.10		4.30			
	MW4	-4.50	0.90				0.90	0.90	0.10		3.50	1.30		
	MA	-3.65	0.90	0.90	0.90	0.90	0.90		0.10		1.50	1.50	1.50	
	MY	-3.30	0.90	0.90	0.90	0.90	0.90	0.90	0.10		-0.50	0.30	0.30	0.10
DAG I	MW2	-1.75	0.90	0.90			0.90	0.90	0.10					
	MW3	-5.95	0.90		0.90		0.90	0.90	0.10		4.30			
	MW4	-4.80	0.90			0.90	0.90	0.90	0.10		3.50	1.30		
	MA	-3.80	0.90	0.90	0.90	0.90	0.90	0.90	0.10		1.50	1.50	1.50	
	MY	-3.20	0.90	0.90	0.90	0.90	0.90	0.90			-0.60	0.10	0.10	0.20
DAG J	MW2	-1.70	0.90	0.90			0.90	0.90	0.10					
	MW3	-5.95	0.90		0.90		0.90	0.90	0.10		4.30			
	MW4	-4.85	0.90			0.90	0.90	0.90	0.10		3.50	1.30		
	MA	-3.80	0.90	0.90	0.90	0.90	0.90	0.90	0.10		1.50	1.50	1.50	
	MY	-3.35	0.90	0.90	0.90	0.90	0.90	0.90	0.10		0.05	0.05	0.05	0.05
*For the complex and positivity violation scenarios models for A and Y also included interactions as follows														
			W1W3	W1W4	W1W5	W3W4	W3W5	W4W5	W1W3W4	W1W3W5	W1W4W5	W3W4W5	W1W3W4W5	
		Positivity violation 1 (30%)												
Complex 1 scenario	A	-2.90	1.30	0.30	1.20	-2.00	-1.90	2.00						
	Y	-0.70	-0.50	1.00	0.10	0.10	0.40	-0.10	-1.20	-1.00	-0.10	-0.40	1.70	
Complex 2 scenario	A	-2.90	1.30	0.30	1.20	-2.00	-1.90	2.00						
	Y	-0.70	-0.90	2.00	0.10	0.20	0.70	-0.20	-2.40	-2.00	-0.30	-0.80	3.40	
		Positivity violation 2 (40%)												
Complex 1 scenario	A	-3.40	2.60	0.70	2.40	-4.00	-3.90	4.00						
	Y	-0.70	-0.50	1.00	0.10	0.10	0.40	-0.10	-1.20	-1.00	-0.10	-0.40	1.70	
Complex 2 scenario	A	-3.40	2.60	0.70	2.40	-4.00	-3.90	4.00						
	Y	-0.70	-0.90	2.00	0.10	0.20	0.70	-0.20	-2.40	-2.00	-0.30	-0.80	3.40	

Table B.1: Coefficient values are displayed for data generation and also for missing indicators across all considered scenarios for positivity violation data

	Model for	Regression coefficient of												
		Intercept	W1	W2	W3	W4	W5	A	Y	B	MW2	MW3	MW4	MZA
Complete Data	W1	-0.40												
	W2	-0.90								-0.50				
	W3	0.40								0.70				
	W4	3.00								-0.10				
	W5	1.00												
		Simple scenario												
	A	-3.85	0.40	0.80	0.50	0.20	0.40			0.70				
	Y	-0.60	-0.30	-0.40	0.20	0.20	0.10	0.20						
		Complex scenarios												
	A		0.40	0.80	0.50	0.20	0.40			0.70				
	Y		-0.30	-0.40	0.20	0.20	0.10	0.20						
DAG T	MW2	-1.10												
	MW3	-2.45									5.60			
	MW4	-4.10									3.90	2.90		
	MA	-1.35									-0.50	1.00	1.00	
	MY	-2.50									0.90	0.90	0.50	0.60
DAG A	MW2	-1.50	0.90				0.10							
	MW3	-2.45	0.90				0.10				3.70			
	MW4	-3.65	0.90				0.10				2.50	2.50		
	MA	-2.55	0.90				0.10				1.30	1.30	1.30	
	MY	-2.25	0.90				0.10				0.30	0.30	0.30	0.35
DAG B	MW2	-1.70	0.90				0.10	0.90						
	MW3	-2.75	0.90				0.10	0.90			4.80			
	MW4	-4.35	0.90				0.10	0.90			3.30	2.70		
	MA	-3.15	0.90	0.90	0.90	0.10	0.10				1.10	1.10	1.00	
	MY	-3.20	0.90	0.90	0.90	0.10	0.10	0.90			0.20	0.20	0.10	0.40
DAG C	MW2	-1.70	0.90				0.10	0.90	0.10					
	MW3	-2.85	0.90				0.10	0.90	0.10		5.00			
	MW4	-4.30	0.90				0.10	0.90	0.10		3.20	2.60		
	MA	-3.20	0.90	0.90	0.90	0.10	0.10		0.10		1.10	1.10	1.10	
	MY	-3.10	0.90	0.90	0.90	0.10	0.10	0.90			0.10	0.10	0.10	0.40
DAG D	MW2	-1.80	0.90	0.90			0.10							
	MW3	-2.85	0.90		0.90		0.10				3.35			
	MW4	-3.90	0.90			0.10	0.10				3.00	2.50		
	MA	-2.95	0.90				0.10	0.90			1.60	1.60	1.50	
	MY	-2.20	0.90				0.10				0.30	0.30	0.20	0.30
DAG E	MW2	-2.00	0.90	0.90			0.10	0.90						
	MW3	-3.55	0.90		0.90		0.10	0.90			5.55			
	MW4	-4.20	0.90			0.10	0.10	0.90			3.30	2.40		
	MA	-3.10	0.90	0.90	0.90	0.10	0.10	0.90			0.80	0.80	0.90	
	MY	-3.05	0.90	0.90	0.90	0.10	0.10	0.90			0.10	0.05	0.05	0.30
DAG F	MW2	-1.80	0.90	0.90			0.10		0.10					
	MW3	-2.40	0.90		0.90		0.10		0.10		3.35			
	MW4	-3.90	0.90			0.10	0.10		0.10		3.00	2.60		
	MA	-2.85	0.90				0.10	0.90	0.10		1.50	1.40	1.40	
	MY	-2.20	0.90				0.10				0.30	0.30	0.20	0.30
DAG G	MW2	-1.55	0.90				0.10	0.90						
	MW3	-2.85	0.90				0.10	0.90			5.20			
	MW4	-4.15	0.90				0.10	0.90			3.30	2.30		
	MA	-3.30	0.90	0.90	0.90	0.10	0.10				1.20	1.20	1.20	
	MY	-3.15	0.90	0.90	0.90	0.10	0.10	0.90	0.10		0.05	0.05	0.05	0.60
DAG H	MW2	-1.70	0.90				0.10	0.90	0.10					
	MW3	-2.90	0.90				0.10	0.90	0.10		5.20			
	MW4	-4.15	0.90				0.10	0.90	0.10		3.30	2.30		
	MA	-3.10	0.90	0.90	0.90	0.10	0.10		0.10		1.00	1.00	1.00	
	MY	-3.10	0.90	0.90	0.90	0.10	0.10	0.90	0.10		0.05	0.05	0.05	0.45
DAG I	MW2	-2.00	0.90	0.90			0.10	0.90	0.10					
	MW3	-3.40	0.90		0.90		0.10	0.90	0.10		5.20			
	MW4	-4.20	0.90			0.10	0.10	0.90	0.10		3.30	2.30		
	MA	-3.25	0.90	0.90	0.90	0.10	0.10	0.90	0.10		1.00	1.00	1.00	
	MY	-3.10	0.90	0.90	0.90	0.10	0.10	0.90			0.05	0.05	0.05	0.40
DAG J	MW2	-2.00	0.90	0.90			0.10	0.90	0.10					
	MW3	-3.40	0.90		0.90		0.10	0.90	0.10		5.20			
	MW4	-4.05	0.90			0.10	0.10	0.90	0.10		3.30	2.30		
	MA	-3.25	0.90	0.90	0.90	0.10	0.10	0.90	0.10		1.00	1.00	1.00	
	MY	-3.10	0.90	0.90	0.90	0.10	0.10	0.90	0.10		0.05	0.05	0.05	0.40

*For the complex scenarios models for A and Y also included interactions as follows

		W1W3	W1W4	W1W5	W3W4	W3W5	W4W5	W1W2W4	W1W2W5	W1W4W5	W2W4W5	W1W2W4W5
		Complex 1 scenario										
	A	-2.95	-0.70	0.50	0.40	-0.50	0.40	-0.55				
	Y	0.20	-0.30	0.20	-0.10	-0.40	0.10	-0.10	-0.10	0.10	-0.10	0.10	0.10	
		Complex 2 scenario										
	A	-3.30	-1.40	1.00	0.80	-1.00	0.80	-1.00				
	Y	1.05	-0.60	0.40	-0.20	-0.80	0.20	-0.20	-0.20	0.20	-0.20	0.20	0.20	

Table B.2: Coefficient values are displayed for data generation and also for missing indicators across all considered scenarios for the modified data

	Model for	Regression coefficient of												
		Intercept	W1	W2	W3	W4	W5	A	Y	B	MW2	MW3	MW4	MZA
Complete Data	W1	0.40												
	W2	0.85								0.30				
	W3	-0.35								0.30				
	W4	3.00								-0.10				
	W5	1.00												
	Simple scenario													
	A	-4.00	0.40	0.80	0.50	0.20	0.40			0.70				
	Y	-0.60	-0.30	-0.40	0.20	0.20	0.10	0.20						
	Complex scenarios													
	A		0.40	0.80	0.50	0.20	0.40			0.70				
	Y		-0.30	-0.40	0.20	0.20	0.10	0.20						
DAG T	MW2	-1.10												
	MW3	-2.45									5.60			
	MW4	-4.10									3.90	2.90		
	MA	-1.35									-0.50	1.00	1.00	
	MY	-2.50									0.90	0.90	0.50	0.60
DAG A	MW2	-1.50	0.90				0.10							
	MW3	-2.45	0.90				0.10				3.70			
	MW4	-3.65	0.90				0.10				2.50	2.50		
	MA	-2.55	0.90				0.10				1.30	1.30	1.30	
	MY	-2.25	0.90				0.10				0.30	0.30	0.30	0.35
DAG B	MW2	-1.70	0.90				0.10	0.90						
	MW3	-2.75	0.90				0.10	0.90			4.80			
	MW4	-4.35	0.90				0.10	0.90			3.30	2.70		
	MA	-3.25	0.90	0.90	0.90	0.10	0.10				1.10	1.10	1.00	
	MY	-3.15	0.90	0.90	0.90	0.10	0.10	0.90			0.20	0.20	0.10	0.10
DAG C	MW2	-1.70	0.90				0.10	0.90	0.10					
	MW3	-2.85	0.90				0.10	0.90	0.10		5.00			
	MW4	-4.30	0.90				0.10	0.90	0.10		3.20	2.60		
	MA	-3.25	0.90	0.90	0.90	0.10	0.10		0.10		1.10	1.10	1.10	
	MY	-3.20	0.90	0.90	0.90	0.10	0.10	0.90			0.15	0.15	0.15	0.25
DAG D	MW2	-1.85	0.90	0.90			0.10							
	MW3	-2.80	0.90		0.90		0.10				3.20			
	MW4	-3.90	0.90			0.10	0.10				3.00	2.60		
	MA	-3.00	0.90				0.10	0.90			1.60	1.60	1.50	
	MY	-2.20	0.90				0.10				0.30	0.30	0.20	0.30
DAG E	MW2	-2.00	0.90	0.90			0.10	0.90						
	MW3	-3.55	0.90		0.90		0.10	0.90			5.55			
	MW4	-4.20	0.90			0.10	0.10	0.90			3.30	2.40		
	MA	-3.15	0.90	0.90	0.90	0.10	0.10	0.90			0.80	0.80	0.90	
	MY	-3.05	0.90	0.90	0.90	0.10	0.10	0.90			0.10	0.05	0.05	0.10
DAG F	MW2	-1.85	0.90	0.90			0.10		0.10					
	MW3	-2.30	0.90		0.90		0.10		0.10		3.20			
	MW4	-3.90	0.90			0.10	0.10		0.10		3.00	2.60		
	MA	-2.85	0.90				0.10	0.90	0.10		1.50	1.40	1.40	
	MY	-2.20	0.90				0.10				0.30	0.30	0.20	0.30
DAG G	MW2	-1.55	0.90				0.10	0.90						
	MW3	-2.95	0.90				0.10	0.90			5.00			
	MW4	-4.05	0.90				0.10	0.90			3.20	2.30		
	MA	-3.35	0.90	0.90	0.90	0.10	0.10				1.20	1.20	1.30	
	MY	-3.15	0.90	0.90	0.90	0.10	0.10	0.90	0.10		0.05	0.05	0.05	0.40
DAG H	MW2	-1.70	0.90				0.10	0.90	0.10					
	MW3	-2.90	0.90				0.10	0.90	0.10		5.20			
	MW4	-4.15	0.90				0.10	0.90	0.10		3.30	2.30		
	MA	-3.15	0.90	0.90	0.90	0.10	0.10		0.10		1.00	1.00	1.00	
	MY	-3.15	0.90	0.90	0.90	0.10	0.10	0.90	0.10		0.05	0.05	0.05	0.40
DAG I	MW2	-2.00	0.90	0.90			0.10	0.90	0.10					
	MW3	-3.45	0.90		0.90		0.10	0.90	0.10		5.10			
	MW4	-4.10	0.90			0.10	0.10	0.90	0.10		3.20	2.30		
	MA	-3.30	0.90	0.90	0.90	0.10	0.10	0.90	0.10		1.00	1.00	1.00	
	MY	-3.05	0.90	0.90	0.90	0.10	0.10	0.90			0.10	0.05	0.05	0.10
DAG J	MW2	-2.00	0.90	0.90			0.10	0.90	0.10					
	MW3	-3.45	0.90		0.90		0.10	0.90	0.10		5.10			
	MW4	-4.10	0.90			0.10	0.10	0.90	0.10		3.20	2.30		
	MA	-3.30	0.90	0.90	0.90	0.10	0.10	0.90	0.10		1.00	1.00	1.00	
	MY	-3.05	0.90	0.90	0.90	0.10	0.10	0.90	0.10		0.05	0.05	0.05	0.15

*For the complex scenarios models for A and Y also included interactions as follows

		W1W3	W1W4	W1W5	W3W4	W3W5	W4W5	W1W2W4	W1W2W5	W1W4W5	W2W4W5	W1W2W4W5
	Complex 1 scenario											
	A	-2.50	-0.60	0.50	0.40	-0.50	0.30	-0.50				
	Y	0.30	-0.30	0.20	-0.10	-0.40	0.10	-0.10	-0.10	0.10	-0.10	0.10	0.10
	Complex 2 scenario											
	A	-2.90	-1.00	1.00	0.80	-0.90	0.50	-0.90				
	Y	1.20	-0.60	0.40	-0.20	-0.80	0.20	-0.20	-0.20	0.20	-0.20	0.20	0.20

Table B.3: Coefficient values are displayed for data generation and also for missing indicators across all considered scenarios for copula 1 data

	Model for	Regression coefficient of												
		Intercept	W1	W2	W3_1	W3_2	W3_3	W4	W5	A	Y	B		
Complete Data	W1	0.40												
	W2	0.90										-0.50		
	W3_1	-0.20										0.70		
	W3_2	-0.30										-1.00		
	W3_3	0.70										-0.30		
	W3_4	-0.80										-0.60		
	W4	3.00										-0.10		
	W5	1.00												
	Simple scenario													
	A	-4.10	0.40	0.80	0.50	0.30	0.60	0.20	0.40			0.70		
	Y	-0.50	-0.30	-0.40	0.20	-0.30	0.35	0.20	0.10	0.20				
	Complex scenarios													
	A		0.40	0.80	0.50	0.30	0.60	0.20	0.40			0.70		
	Y		-0.30	-0.40	0.20	-0.30	0.35	0.20	0.10	0.20				
*For the complex scenarios models for A and Y also included interactions as follows														
			W1W3_1	W1W3_2	W1W3_3	W1W4	W1W5	W3_1W4	W3_2W4	W3_3W4	W3_1W5	W3_2W5	W3_3W5	W4W5
	Complex 1 scenario													
	A	-2.40	-0.40	0.30	0.30	0.30	-0.20	-0.30	-0.50	0.10	-0.20	-0.50	-0.40	-0.30
	Y	0.15	-0.30	0.40	-0.50	0.20	-0.10	-0.40	0.20	-0.40	0.10	-0.30	-0.20	-0.10
	Complex 2 scenario													
	A	-2.55	-0.80	0.60	0.60	0.50	-0.40	-0.50	-0.80	0.20	-0.40	-0.90	-0.70	-0.50
	Y	0.60	-0.60	0.70	-0.90	0.40	-0.20	-0.70	0.40	-0.70	0.20	-0.60	-0.40	-0.20

		W1W2W4	W1W2W5	W1W4W5	W2W4W5	W1W2W4W5
	Complex 1 scenario					
	A					
	Y	-0.10	0.10	-0.10	0.10	0.10
	Complex 2 scenario					
	A					
	Y	-0.20	0.20	-0.20	0.20	0.20

Table B.4: Coefficient values are displayed for data generation across all considered scenarios for copula 2 data

	Model for	Regression coefficient of											
		Intercept	W1	W2	W3_1	W3_2	W3_3	W4	W5	W6	A	Y	B
Complete Data	W1	0.40											
	W2	0.90											-0.50
	W3_1	-0.20											0.70
	W3_2	-0.30											-1.00
	W3_3	0.70											-0.30
	W3_4	-0.80											-0.60
	W4	3.00											-0.10
	W5	1.00											
	W6 shape	2											0.10
	W6 rate	1											0.10
	Simple scenario												
	A	-4.90	0.40	0.80	0.50	0.30	0.60	0.20	0.40	0.40			0.70
	Y	-0.05	-0.30	-0.40	0.20	-0.30	0.35	0.20	0.10	-0.20	0.20		
	Complex scenarios												
	A		0.40	0.80	0.50	0.30	0.60	0.20	0.40	0.40			0.70
	Y		-0.30	-0.40	0.20	-0.30	0.35	0.20	0.10	-0.20	0.20		

*For the complex scenarios models for A and Y also included interactions as follows

			W1W3_1	W1W3_2	W1W3_3	W1W4	W1W5	W3_1W4	W3_2W4	W3_3W4	W3_1W5	W3_2W5	W3_3W5	W4W5
	Complex 1 scenario													
	A	-2.80	-0.60	0.40	0.50	0.60	-0.60	-0.50	0.60	0.20	-0.40	-0.60	-0.30	-0.30
	Y	0.60	-0.30	0.40	-0.50	0.20	-0.10	-0.40	0.20	-0.40	0.10	-0.30	-0.20	-0.10
	Complex 2 scenario													
	A	-2.85	-1.10	0.80	0.90	1.10	-1.10	-0.90	1.10	0.40	-0.70	-1.10	-0.50	-0.50
	Y	1.30	-0.60	0.80	-1.00	0.40	-0.20	-0.80	0.40	-0.80	0.20	-0.60	-0.40	-0.20

		W1W6	W4W6	W5W6	W1W4W6	W1W5W6	W1W4W5	W4W5W6	W1W4W5W6
Complex 1 scenario									
A		-0.40	-0.30	0.20					
Y		-0.20	-0.10	0.10	-0.10	0.10	-0.10	0.10	0.10
Complex 2 scenario									
A		-0.70	-0.50	0.40					
Y		-0.40	-0.20	0.20	-0.20	0.20	-0.20	0.20	0.20

Table B.5: Coefficient values are displayed for data generation across all considered scenarios for copula 3 data

	Model for	Regression coefficient of												
		Intercept	W1	W2	W3_1	W3_3	W4	W5	A	Y	B	MW2	MW3	MW4	MZA
	MW2	-1.10													
	MW3	-2.45										5.60			
DAG T	MW4	-4.15										3.90	2.90		
	MA	-1.35										-0.50	1.00	1.00	
	MY	-2.50										0.90	0.90	0.50	0.60
	MW2	-1.50	0.90					0.10							
	MW3	-2.45	0.90					0.10				3.70			
DAG A	MW4	-3.65	0.90					0.10				2.50	2.50		
	MA	-2.55	0.90					0.10				1.30	1.30	1.30	
	MY	-2.25	0.90					0.10				0.30	0.30	0.30	0.35
	MW2	-1.65	0.90					0.10	0.90						
	MW3	-2.70	0.90					0.10	0.90			4.20			
DAG B	MW4	-4.35	0.90					0.10	0.90			3.10	2.70		
	MA	-3.10	0.90	0.90	0.90	0.90	0.10	0.10				1.10	1.10	1.10	
	MY	-3.10	0.90	0.90	0.90	0.90	0.10	0.10	0.90			0.20	0.20	0.10	0.35
	MW2	-1.65	0.90					0.10	0.90	0.10					
	MW3	-2.75	0.90					0.10	0.90	0.10		4.50			
DAG C	MW4	-4.30	0.90					0.10	0.90	0.10		3.20	2.60		
	MA	-3.10	0.90	0.90	0.90	0.90	0.10	0.10		0.10		1.10	1.10	1.10	
	MY	-3.05	0.90	0.90	0.90	0.90	0.10	0.10	0.90			0.10	0.10	0.10	0.40
	MW2	-1.85	0.90	0.90				0.10							
	MW3	-2.75	0.90		0.90	0.90		0.10				3.30			
DAG D	MW4	-4.20	0.90				0.10	0.10				3.00	2.95		
	MA	-2.95	0.90					0.10	0.90			1.60	1.60	1.50	
	MY	-2.20	0.90					0.10				0.30	0.30	0.20	0.30
	MW2	-2.00	0.90	0.90				0.10	0.90						
	MW3	-3.50	0.90		0.90	0.90		0.10	0.90			5.55			
DAG E	MW4	-4.20	0.90				0.10	0.10	0.90			3.30	2.40		
	MA	-3.00	0.90	0.90	0.90	0.90	0.10	0.10	0.90			0.80	0.80	0.90	
	MY	-2.95	0.90	0.90	0.90	0.90	0.10	0.10	0.90			0.10	0.05	0.05	0.25
	MW2	-1.85	0.90	0.90				0.10		0.10					
	MW3	-2.75	0.90		0.90	0.90		0.10		0.10		3.35			
DAG F	MW4	-4.00	0.90				0.10	0.10		0.10		3.00	2.60		
	MA	-3.00	0.90					0.10	0.90	0.10		1.60	1.60	1.60	
	MY	-2.20	0.90					0.10				0.30	0.30	0.20	0.30
	MW2	-1.65	0.90					0.10	0.90						
	MW3	-3.05	0.90					0.10	0.90			5.90			
DAG G	MW4	-4.35	0.90					0.10	0.90			3.30	2.70		
	MA	-2.95	0.90	0.90	0.90	0.90	0.10	0.10				0.90	0.90	1.00	
	MY	-3.00	0.90	0.90	0.90	0.90	0.10	0.10	0.90	0.10		0.05	0.05	0.05	0.40
	MW2	-1.70	0.90					0.10	0.90	0.10					
	MW3	-3.05	0.90					0.10	0.90	0.10		5.90			
DAG H	MW4	-4.35	0.90					0.10	0.90	0.10		3.30	2.70		
	MA	-2.95	0.90	0.90	0.90	0.90	0.10	0.10		0.10		1.00	0.90	0.90	
	MY	-3.00	0.90	0.90	0.90	0.90	0.10	0.10	0.90	0.10		0.10	0.10	0.10	0.35
	MW2	-2.00	0.90	0.90				0.10	0.90	0.10					
	MW3	-3.45	0.90		0.90	0.90		0.10	0.90	0.10		5.50			
DAG I	MW4	-4.20	0.90				0.10	0.10	0.90	0.10		3.30	2.35		
	MA	-2.95	0.90	0.90	0.90	0.90	0.10	0.10	0.90	0.10		0.80	0.80	0.80	
	MY	-2.95	0.90	0.90	0.90	0.90	0.10	0.10	0.90			0.05	0.05	0.05	0.25
	MW2	-2.00	0.90	0.90				0.10	0.90	0.10					
	MW3	-3.45	0.90		0.90	0.90		0.10	0.90	0.10		5.50			
DAG J	MW4	-4.20	0.90				0.10	0.10	0.90	0.10		3.30	2.35		
	MA	-2.95	0.90	0.90	0.90	0.90	0.10	0.10	0.90	0.10		0.80	0.80	0.80	
	MY	-2.95	0.90	0.90	0.90	0.90	0.10	0.10	0.90	0.10		0.05	0.05	0.05	0.25

Table B.6: Shows coefficient values for missing indicators across all considered scenarios for copula 2 data

	Model for	Regression coefficient of														
		Intercept	W1	W2	W3_1	W3_3	W4	W5	W6	A	Y	B	MW2	MW3	MW4	MW6	MZA
DAG T	MW2	-1.10															
	MW3	-2.45											5.60				
	MW4	-4.10											3.90	2.80			
	MW6	-4.40											1.90	1.90	1.90		
	MA	-1.35											-0.50	0.70	0.70	0.80	
	MY	-2.60											0.90	0.90	0.50	0.50	0.40
DAG A	MW2	-1.50	0.90					0.10									
	MW3	-2.45	0.90					0.10					3.70				
	MW4	-3.65	0.90					0.10					2.50	2.50			
	MW6	-5.15	0.90					0.10					1.90	2.30	2.20		
	MA	-2.55	0.90					0.10					1.10	1.10	1.00	1.00	
	MY	-2.25	0.90					0.10					0.30	0.30	0.20	0.20	0.20
DAG B	MW2	-1.70	0.90					0.10		0.90							
	MW3	-2.95	0.90					0.10		0.90			5.40				
	MW4	-4.30	0.90					0.10		0.90			3.30	2.50			
	MW6	-5.95	0.90					0.10		0.90			3.20	1.80	1.90		
	MA	-3.10	0.90	0.90	0.90	0.90	0.10	0.10	0.10				0.90	0.90	0.90	0.90	
	MY	-3.00	0.90	0.90	0.90	0.90	0.10	0.10	0.10	0.90			0.10	0.10	0.10	0.20	0.10
DAG C	MW2	-1.70	0.90					0.10		0.90	0.10						
	MW3	-2.95	0.90					0.10		0.90	0.10		5.40				
	MW4	-4.35	0.90					0.10		0.90	0.10		3.30	2.50			
	MW6	-6.05	0.90					0.10		0.90	0.10		3.20	1.80	1.90		
	MA	-3.10	0.90	0.90	0.90	0.90	0.10	0.10	0.10		0.10		0.80	0.90	0.90	0.90	
	MY	-3.05	0.90	0.90	0.90	0.90	0.10	0.10	0.10	0.90			0.10	0.10	0.10	0.20	0.25
DAG D	MW2	-1.85	0.90	0.90				0.10									
	MW3	-2.75	0.90		0.90	0.90		0.10					3.35				
	MW4	-4.00	0.90				0.10	0.10					3.00	2.70			
	MW6	-5.30	0.90					0.10	0.10				2.20	2.20	2.30		
	MA	-2.85	0.90					0.10		0.90			1.30	1.20	1.20	1.20	
	MY	-2.20	0.90					0.10					0.30	0.30	0.20	0.20	0.10
DAG E	MW2	-2.00	0.90	0.90				0.10		0.90							
	MW3	-3.50	0.90		0.90	0.90		0.10		0.90			5.50				
	MW4	-4.15	0.90				0.10	0.10		0.90			3.30	2.20			
	MW6	-5.90	0.90					0.10	0.10	0.90			3.20	1.70	1.90		
	MA	-2.95	0.90	0.90	0.90	0.90	0.10	0.10	0.10	0.90			0.70	0.60	0.60	0.70	
	MY	-3.00	0.90	0.90	0.90	0.90	0.10	0.10	0.10	0.90			0.10	0.05	0.05	0.05	0.25
DAG F	MW2	-1.85	0.90	0.90				0.10			0.10						
	MW3	-2.75	0.90		0.90	0.90		0.10			0.10		3.30				
	MW4	-4.00	0.90				0.10	0.10			0.10		3.00	2.70			
	MW6	-5.40	0.90					0.10	0.10		0.10		2.20	2.30	2.20		
	MA	-2.90	0.90					0.10		0.90	0.10		1.20	1.20	1.20	1.20	
	MY	-2.20	0.90					0.10					0.30	0.30	0.20	0.20	0.10
DAG G	MW2	-1.70	0.90					0.10		0.90							
	MW3	-2.95	0.90					0.10		0.90			5.50				
	MW4	-4.55	0.90					0.10		0.90			3.30	2.70			
	MW6	-6.30	0.90					0.10		0.90			3.20	2.20	2.10		
	MA	-2.95	0.90	0.90	0.90	0.90	0.10	0.10	0.10				0.80	0.80	0.80	0.60	
	MY	-3.05	0.90	0.90	0.90	0.90	0.10	0.10	0.10	0.90	0.10		0.05	0.05	0.05	0.05	0.40
DAG H	MW2	-1.70	0.90					0.10		0.90	0.10						
	MW3	-3.05	0.90					0.10		0.90	0.10		5.50				
	MW4	-4.35	0.90					0.10		0.90	0.10		3.30	2.50			
	MW6	-6.05	0.90					0.10		0.90	0.10		3.20	1.80	1.90		
	MA	-2.95	0.90	0.90	0.90	0.90	0.10	0.10	0.10		0.10		0.80	0.80	0.80	0.60	
	MY	-3.00	0.90	0.90	0.90	0.90	0.10	0.10	0.10	0.90	0.10		0.05	0.05	0.05	0.05	0.30
DAG I	MW2	-2.00	0.90	0.90				0.10		0.90	0.10						
	MW3	-3.50	0.90		0.90	0.90		0.10		0.90	0.10		5.50				
	MW4	-4.20	0.90				0.10	0.10		0.90	0.10		3.30	2.20			
	MW6	-5.90	0.90					0.10	0.10	0.90	0.10		3.20	1.70	1.90		
	MA	-2.95	0.90	0.90	0.90	0.90	0.10	0.10	0.10	0.90	0.10		0.70	0.60	0.60	0.70	
	MY	-2.95	0.90	0.90	0.90	0.90	0.10	0.10	0.10	0.90			0.10	0.05	0.05	0.05	0.15
DAG J	MW2	-2.00	0.90	0.90				0.10		0.90	0.10						
	MW3	-3.50	0.90		0.90	0.90		0.10		0.90	0.10		5.50				
	MW4	-4.20	0.90				0.10	0.10		0.90	0.10		3.30	2.20			
	MW6	-5.90	0.90					0.10	0.10	0.90	0.10		3.20	1.70	1.90		
	MA	-2.95	0.90	0.90	0.90	0.90	0.10	0.10	0.10	0.90	0.10		0.70	0.60	0.60	0.70	
	MY	-2.95	0.90	0.90	0.90	0.90	0.10	0.10	0.10	0.90	0.10		0.05	0.05	0.05	0.05	0.15

Table B.7: Shows coefficient values for missing indicators across all considered scenarios for copula 3 data

Parameter 1 represents proportion for binary/categorical variables and mean for normal distributed outcome. Parameter 2 represents the standard deviation and for Y it is ordered regarding scenario complexity.

Distribution of variables in simulated complete data

		W1	W2	W3	W4	W5	A	Y
	Type	binary	binary	binary	normal	normal	binary	normal
For all scenarios	Parameter 1	0.40	0.30	0.59	3	1	0.15	0
	Parameter 2				1	2		1;1.3;2.2

		% with missing value						
		W2	W3	W4	A	Y	A/Y	Any
	DAG T	25	30	25	30	20	40	50
	DAG A	25	30	25	30	20	40	50
	DAG B	25	30	25	30	20	40	50
	DAG C	25	30	25	30	20	40	50
	DAG D	25	30	25	30	20	40	50
For all scenarios	DAG E	25	30	25	30	20	40	50
	DAG F	25	30	25	30	20	40	50
	DAG G	25	30	25	30	20	40	50
	DAG H	25	30	25	30	20	40	50
	DAG I	25	30	25	30	20	40	50
	DAG J	25	30	25	30	20	40	50

Table B.8: Distribution of variables in the simulated complete data and proportion with missingness across the 11 simulated missingness mechanisms and the different considered scenarios

Parameter 1 represents proportion for binary/categorical variables and mean for normal distributed outcome. Parameter 2 represents the standard deviation and for Y it is ordered regarding scenario complexity.

Distribution of variables in simulated complete data

		W1	W2	W3_1	W3_2	W3_3	W3_4	W4	W5	A	Y
	Type	binary	binary	categorical	categorical	categorical	categorical	normal	normal	binary	normal
For all scenarios	Parameter 1	0.40	0.30	0.25	0.56	0.20	0.09	3	1	0.15	0
	Parameter 2							1	2		1;1.5;2.7

		% with missing value						
		W2	W3	W4	A	Y	A/Y	Any
	DAG T	25	30	25	30	20	40	50
	DAG A	25	30	25	30	20	40	50
	DAG B	25	30	25	30	20	40	50
	DAG C	25	30	25	30	20	40	50
	DAG D	25	30	25	30	20	40	50
For all scenarios	DAG E	25	30	25	30	20	40	50
	DAG F	25	30	25	30	20	40	50
	DAG G	25	30	25	30	20	40	50
	DAG H	25	30	25	30	20	40	50
	DAG I	25	30	25	30	20	40	50
	DAG J	25	30	25	30	20	40	50

Table B.9: Distribution of variables in the simulated complete data and proportion with missingness across the 11 simulated missingness mechanisms and the different considered scenarios.

Parameter 1 represents proportion for binary/categorical variables, shape for gamma distributed variable and mean for normal distributed outcome. Parameter 2 represents rate for gamma distributed variable and the standard deviation and for Y its ordered regarding scenario complexity.

Distribution of variables in simulated complete data

		W1	W2	W3_1	W3_2	W3_3	W3_4	W4	W5	W6	A	Y
	Type	binary	binary	categorical	categorical	categorical	categorical	normal	normal	gamma	binary	normal
For all scenarios	Parameter 1	0.40	0.30	0.25	0.56	0.20	0.09	3	1	2	0.15	0
	Parameter 2							1	2	1		1;2.5;4.5

% with missing value

		W2	W3	W4	W6	A	Y	A/Y	Any
	DAG T	25	30	25	30	30	20	40	50
	DAG A	25	30	25	30	30	20	40	50
	DAG B	25	30	25	30	30	20	40	50
	DAG C	25	30	25	30	30	20	40	50
	DAG D	25	30	25	30	30	20	40	50
For all scenarios	DAG E	25	30	25	30	30	20	40	50
	DAG F	25	30	25	30	30	20	40	50
	DAG G	25	30	25	30	30	20	40	50
	DAG H	25	30	25	30	30	20	40	50
	DAG I	25	30	25	30	30	20	40	50
	DAG J	25	30	25	30	30	20	40	50

Table B.10: Distribution of variables in the simulated complete data and proportion with missingness across the 11 simulated missingness mechanisms and the different considered scenarios.

Causal diagram	Missing data method	Estimated mean	Bias	Relative Bias (%)	empSE	RMSE	ModSE	Error in ModSE (%)	Coverage	Bias eliminated coverage	Mean CI length	Proportional CI length
A	CC	0.20	-0.00	-0.51	0.11	0.11	0.11	-4.93	0.93	0.93	0.42	1.00
	Ext	0.20	-0.00	-0.33	0.11	0.11	0.11	-5.13	0.93	0.93	0.42	1.00
	Ext MCMI	0.22	0.02	9.98	0.10	0.10	0.10	-5.95	0.92	0.93	0.38	0.90
	MI Reg	0.20	-0.00	-1.09	0.10	0.10	0.10	4.57	0.95	0.95	0.40	0.97
	MI PMM	0.20	-0.00	-1.02	0.10	0.10	0.11	5.41	0.95	0.96	0.41	0.98
	MI Cart	0.19	-0.01	-5.56	0.10	0.10	0.10	5.03	0.94	0.95	0.39	0.94
	MI RF	0.17	-0.03	-13.90	0.08	0.08	0.10	29.64	0.98	0.98	0.39	0.95
	MI Amelia	0.18	-0.02	-7.76	0.07	0.08	0.09	20.83	0.97	0.97	0.35	0.84
	MI 2Int	0.19	-0.01	-3.96	0.10	0.10	0.11	4.98	0.95	0.95	0.41	0.99
	MI 3Int	0.19	-0.01	-4.78	0.10	0.10	0.11	5.04	0.95	0.95	0.41	0.98
	MI 2IntN	0.19	-0.01	-5.01	0.10	0.10	0.11	8.32	0.95	0.95	0.41	0.98
	MI 3IntN	0.19	-0.01	-5.50	0.10	0.10	0.11	7.01	0.95	0.95	0.41	0.97
B	CC	0.20	0.00	1.84	0.15	0.15	0.13	-11.53	0.92	0.92	0.51	1.00
	Ext	0.20	0.00	2.01	0.15	0.15	0.13	-10.89	0.92	0.91	0.51	1.00
	Ext MCMI	0.22	0.02	8.95	0.13	0.13	0.12	-10.90	0.90	0.92	0.45	0.89
	MI Reg	0.19	-0.01	-5.03	0.12	0.12	0.12	0.13	0.93	0.93	0.47	0.92
	MI PMM	0.19	-0.01	-4.95	0.12	0.12	0.13	3.32	0.94	0.94	0.48	0.94
	MI Cart	0.17	-0.03	-13.42	0.11	0.12	0.12	6.47	0.95	0.95	0.46	0.90
	MI RF	0.14	-0.06	-28.39	0.09	0.10	0.12	34.23	0.96	0.98	0.45	0.88
	MI Amelia	0.18	-0.02	-11.68	0.08	0.09	0.10	21.39	0.96	0.97	0.40	0.77
	MI 2Int	0.19	-0.01	-6.18	0.13	0.13	0.13	-1.01	0.93	0.93	0.49	0.96
	MI 3Int	0.19	-0.01	-6.52	0.13	0.13	0.13	-1.17	0.92	0.92	0.48	0.95
	MI 2IntN	0.18	-0.02	-9.92	0.12	0.12	0.13	3.33	0.93	0.93	0.48	0.95
	MI 3IntN	0.18	-0.02	-9.82	0.12	0.12	0.12	2.94	0.94	0.94	0.48	0.94
C	CC	0.18	-0.02	-10.94	0.14	0.14	0.13	-4.65	0.93	0.93	0.52	1.00
	Ext	0.18	-0.02	-11.05	0.14	0.14	0.13	-4.63	0.92	0.93	0.51	0.99
	Ext MCMI	0.19	-0.01	-7.16	0.12	0.12	0.12	-5.41	0.93	0.93	0.45	0.87
	MI Reg	0.18	-0.02	-11.43	0.12	0.12	0.12	6.96	0.94	0.94	0.48	0.92
	MI PMM	0.18	-0.02	-11.83	0.11	0.12	0.13	11.47	0.95	0.95	0.49	0.94
	MI Cart	0.16	-0.04	-19.89	0.11	0.12	0.12	10.01	0.94	0.96	0.46	0.88
	MI RF	0.13	-0.07	-36.11	0.08	0.11	0.12	46.06	0.96	0.99	0.45	0.87
	MI Amelia	0.18	-0.02	-8.87	0.08	0.09	0.10	25.18	0.97	0.97	0.40	0.78
	MI 2Int	0.18	-0.02	-10.20	0.13	0.13	0.13	3.97	0.94	0.94	0.50	0.97
	MI 3Int	0.18	-0.02	-8.66	0.12	0.12	0.13	4.78	0.94	0.94	0.49	0.95
	MI 2IntN	0.17	-0.03	-14.66	0.11	0.12	0.13	10.08	0.95	0.95	0.48	0.93
	MI 3IntN	0.17	-0.03	-13.91	0.11	0.12	0.13	10.51	0.95	0.95	0.48	0.93
D	CC	0.20	0.00	0.65	0.12	0.12	0.12	-1.64	0.94	0.94	0.45	0.99
	Ext	0.20	0.00	0.21	0.12	0.12	0.12	-1.55	0.93	0.93	0.45	1.00
	Ext MCMI	0.23	0.03	12.79	0.11	0.11	0.10	-3.73	0.93	0.93	0.41	0.90
	MI Reg	0.21	0.01	3.33	0.10	0.10	0.11	9.46	0.94	0.94	0.43	0.96
	MI PMM	0.21	0.01	3.56	0.10	0.10	0.11	10.76	0.96	0.96	0.44	0.98
	MI Cart	0.19	-0.01	-3.40	0.10	0.10	0.11	10.80	0.94	0.94	0.42	0.94
	MI RF	0.17	-0.03	-14.57	0.08	0.09	0.11	34.66	0.98	0.99	0.43	0.95
	MI Amelia	0.19	-0.01	-4.00	0.07	0.07	0.10	30.21	0.98	0.98	0.37	0.83
	MI 2Int	0.20	0.00	0.75	0.11	0.11	0.12	10.65	0.95	0.95	0.45	1.00
	MI 3Int	0.20	-0.00	-0.60	0.11	0.11	0.11	8.52	0.96	0.96	0.44	0.98
	MI 2IntN	0.19	-0.01	-3.27	0.10	0.10	0.12	13.82	0.96	0.96	0.44	0.99
	MI 3IntN	0.20	-0.00	-1.90	0.10	0.10	0.12	14.70	0.97	0.97	0.44	0.99
E	CC	0.20	0.00	0.27	0.15	0.15	0.14	-7.33	0.91	0.91	0.54	0.99
	Ext	0.20	0.00	0.11	0.15	0.15	0.14	-7.05	0.92	0.92	0.54	1.00
	Ext MCMI	0.22	0.02	8.71	0.14	0.14	0.12	-9.37	0.91	0.90	0.48	0.88
	MI Reg	0.21	0.01	4.34	0.13	0.13	0.14	5.37	0.94	0.94	0.52	0.96
	MI PMM	0.21	0.01	4.36	0.13	0.13	0.14	6.83	0.95	0.94	0.52	0.96
	MI Cart	0.18	-0.02	-9.41	0.11	0.11	0.13	12.90	0.95	0.96	0.49	0.90
	MI RF	0.15	-0.05	-27.22	0.09	0.11	0.12	37.25	0.98	0.98	0.48	0.88
	MI Amelia	0.19	-0.01	-5.02	0.09	0.09	0.11	25.70	0.98	0.98	0.43	0.78
	MI 2Int	0.21	0.01	4.18	0.13	0.13	0.14	2.25	0.93	0.93	0.53	0.97
	MI 3Int	0.20	0.00	2.44	0.13	0.13	0.14	3.28	0.93	0.93	0.52	0.95
	MI 2IntN	0.20	-0.00	-2.04	0.13	0.13	0.14	8.41	0.94	0.94	0.52	0.96
	MI 3IntN	0.20	-0.00	-2.11	0.12	0.13	0.14	9.38	0.96	0.95	0.52	0.96
F	CC	0.19	-0.01	-4.72	0.12	0.13	0.12	-7.18	0.93	0.93	0.45	0.99
	Ext	0.19	-0.01	-5.00	0.12	0.12	0.12	-6.01	0.93	0.93	0.45	1.00
	Ext MCMI	0.21	0.01	7.47	0.11	0.12	0.10	-8.48	0.93	0.93	0.41	0.90
	MI Reg	0.20	0.00	0.07	0.11	0.11	0.11	2.58	0.94	0.94	0.44	0.98
	MI PMM	0.20	-0.00	-0.13	0.11	0.11	0.11	1.99	0.94	0.94	0.44	0.98
	MI Cart	0.19	-0.01	-6.37	0.11	0.11	0.11	3.06	0.94	0.94	0.43	0.94
	MI RF	0.16	-0.04	-18.75	0.09	0.09	0.11	27.85	0.97	0.99	0.42	0.94
	MI Amelia	0.20	-0.00	-0.61	0.08	0.08	0.10	23.30	0.98	0.98	0.38	0.84
	MI 2Int	0.20	-0.00	-1.12	0.12	0.12	0.12	1.14	0.93	0.93	0.45	1.00
	MI 3Int	0.20	-0.00	-1.60	0.11	0.11	0.11	1.44	0.95	0.94	0.44	0.98
	MI 2IntN	0.19	-0.01	-7.16	0.11	0.11	0.12	7.45	0.95	0.95	0.45	0.98
	MI 3IntN	0.19	-0.01	-6.69	0.11	0.11	0.11	5.35	0.94	0.94	0.44	0.97
G	CC	0.19	-0.01	-6.00	0.14	0.14	0.13	-7.43	0.92	0.92	0.51	1.00
	Ext	0.19	-0.01	-4.94	0.14	0.14	0.13	-7.41	0.92	0.92	0.51	0.99
	Ext MCMI	0.20	0.00	1.88	0.12	0.12	0.12	-7.61	0.92	0.92	0.45	0.87
	MI Reg	0.17	-0.03	-12.74	0.12	0.12	0.13	7.16	0.95	0.95	0.48	0.94
	MI PMM	0.18	-0.02	-11.51	0.12	0.12	0.13	6.72	0.94	0.95	0.49	0.95
	MI Cart	0.16	-0.04	-18.66	0.11	0.12	0.12	6.84	0.93	0.94	0.45	0.89
	MI RF	0.14	-0.06	-32.03	0.08	0.11	0.12	38.29	0.95	0.98	0.45	0.88
	MI Amelia	0.17	-0.03	-16.82	0.08	0.09	0.10	27.21	0.96	0.97	0.40	0.78
	MI 2Int	0.17	-0.03	-13.56	0.13	0.13	0.13	1.11	0.93	0.94	0.49	0.96
	MI 3Int	0.17	-0.03	-14.18	0.12	0.13	0.12	0.09	0.93	0.94	0.48	0.93
	MI 2IntN	0.17	-0.03	-16.96	0.12	0.12	0.12	6.18	0.93	0.94	0.48	0.93
	MI 3IntN	0.17	-0.03	-16.28	0.12	0.12	0.12	4.28	0.93	0.94	0.48	0.93
H	CC	0.17	-0.03	-17.48	0.14	0.15	0.13	-7.60	0.91	0.91	0.51	1.00
	Ext	0.17	-0.03	-17.40	0.14	0.15	0.13	-8.11	0.91	0.92	0.51	0.99
	Ext MCMI	0.17	-0.03	-12.87	0.13	0.13	0.11	-8.74	0.91	0.93	0.45	0.87
	MI Reg	0.16	-0.04	-19.16	0.12	0.12	0.12	4.98	0.93	0.95	0.48	0.93
	MI PMM	0.16	-0.04	-19.65	0.12	0.12	0.13	5.55	0.93	0.95	0.48	0.94
	MI Cart	0.15	-0.05	-26.27	0.11	0.12	0.12	6.01	0.93	0.95	0.45	0.88
	MI RF	0.12	-0.08	-40.33	0.08	0.12	0.12	39.42	0.94	0.99	0.45	0.87
	MI Amelia	0.17	-0.03	-14.10	0.08	0.09	0.10	24.90	0.96	0.97	0.40	0.78
	MI 2Int	0.17	-0.03	-16.87	0.13	0.13	0.13	0.19	0.91	0.93	0.49	0.96
	MI 3Int	0.17	-0.03	-17.18	0.12	0.13	0.13	2.26	0.92	0.94	0.48	0.95
	MI 2IntN	0.16	-0.04	-21.91	0.12	0.12	0.12	6.80	0.92	0.94	0.48	0.93
	MI 3IntN	0.16	-0.04	-21.67	0.12	0.13	0.12	6.08	0.92	0.94	0.48	0.93
I	CC	0.16	-0.04	-18.68	0.15	0.16	0.14	-9.06	0.90	0.91	0.53	0.99
	Ext	0.16	-0.04	-20.02	0.15	0.16	0.14	-7.91	0.90	0.91	0.54	0.99
	Ext MCMI	0.17	-0.03	-16.20	0.13	0.14	0.12	-9.21	0.91	0.92	0.47	0.88
	MI Reg	0.18	-0.02	-9.49	0.13	0.13	0.13	6.47	0.94	0.95	0.52	0.96
	MI PMM	0.18	-0.02	-8.58	0.13	0.13	0.14	8.17	0.94	0.95	0.53	0.98
	MI Cart	0.16	-0.04	-19.98	0.12	0.12	0.13	8.47	0.94	0.95	0.49	0.91
	MI RF	0.12	-0.08	-40.74	0.09	0.12	0.13	42.82	0.96	0.99	0.49	0.90
	MI Amelia	0.18	-0.02	-8.25	0.09	0.09	0.11	23.40	0.97	0.97	0.42	0.78
	MI 2Int	0.19	-0.01	-4.92	0.14	0.14	0.14	1.50	0.92	0.92	0.54	1.00
	MI 3Int	0.19	-0.01	-4.57	0.14	0.14	0.14	0.02	0.92	0.93	0.52	0.97
	MI 2IntN	0.17	-0.03	-13.68	0.13	0.13	0.14	4.89	0.94	0.94	0.52	0.97
	MI 3IntN	0.17	-0.03	-13.21	0.12	0.13	0.14	10.13	0.94	0.95	0.52	0.97
J	CC	0.16	-0.04	-21.47	0.15	0.16	0.13	-13.19	0.90	0.92	0.52	0.98
	Ext	0.16	-0.04	-22.44	0.15	0.16	0.14	-10.61	0.90	0.90	0.52	0.99
	Ext MCMI	0.16	-0.04	-19.38	0.14	0.14	0.12	-13.11	0.89	0.91	0.46	0.87
	MI Reg	0.17	-0.03	-15.51	0.13	0.13	0.13	4.73	0.93	0.94	0.51	0.97
	MI PMM	0.17	-0.03	-14.28	0.13	0.13	0.14	3.79	0.93	0.94	0.52	0.99
	MI Cart	0.15	-0.05	-24.29	0.12	0.13	0.12	3.87	0.93	0.95	0.48	0.91
	MI RF	0.11	-0.09	-42.82	0.09	0.12	0.12	36.44	0.95	0.99	0.48	0.90
	MI Amelia	0.17	-0.03	-12.74	0.09	0.09	0.11	23.35	0.97	0.98	0.42	0.79
	MI 2Int	0.18	-0.02	-10.52	0.14	0.14	0.14	1.54	0.93	0.93	0.53	1.00
	MI 3Int	0.18	-0.02	-10.54	0.13	0.13	0.14	3.54	0.93	0.93	0.52	0.98
	MI 2IntN	0.16	-0.04	-19.49	0.13	0.13	0.13	6.78	0.94	0.94	0.51	0.97
	MI 3IntN	0.16	-0.04	-17.97	0.12	0.13	0.13	7.21	0.94	0.94	0.51	0.96
T	CC	0.20	0.00	0.16	0.11	0.11	0.11	-5.38	0.92	0.92	0.41	0.99
	Ext	0.20	-0.00	-0.35	0.11	0.11	0.11	-4.79	0.92	0.92	0.42	1.00
	Ext MCMI	0.22	0.02	11.02	0.10	0.10	0.10	-4.02	0.93	0.94	0.37	0.90
	MI Reg	0.20	0.00	0.63	0.10	0.10	0.10	8.62	0.95	0.95	0.40	0.97
	MI PMM	0.20	-0.00	-0.39	0.10	0.10	0.11	7.98	0.95	0.95	0.41	0.99
	MI Cart	0.19	-0.01	-5.01	0.10	0.10	0.10	5.54	0.95	0.94	0.39	0.95
	MI RF	0.17	-0.03	-13.39	0.07	0.08	0.10	33.92	0.98	0.99	0.39	0.93
	MI Amelia	0.19	-0.01	-6.70	0.07	0.07	0.09	29.23	0.98	0.98	0.36	0.86
	MI 2Int	0.19	-0.01	-3.64	0.10	0.10	0.11	8.99	0.95	0.95	0.42	1.00
	MI 3Int	0.19	-0.01	-3.90	0.10	0.10	0.11	10.44	0.96	0.96	0.41	0.99
	MI 2IntN	0.19	-0.01	-6.49	0.09	0.10	0.11	12.23	0.95	0.95	0.41	0.99
	MI 3IntN	0.19	-0.01	-6.08	0.09	0.09	0.11	12.38	0.96	0.96	0.41	0.98

Table B.11: Performance of missing data methods with *TMLE* implementation for the 11 assessed missingness mechanisms under the simple scenario for reproduced data

Causal diagram	Missing data method	Estimated mean	Bias	Relative Bias (%)	empSE	RMSE	ModSE	Error in ModSE (%)	Coverage	Bias eliminated coverage	Mean CI length	Proportional CI length
A	CC	0.21	0.01	5.45	0.11	0.11	0.10	-13.73	0.90	0.90	0.38	0.90
	Ext	0.21	0.01	5.04	0.12	0.12	0.10	-14.44	0.89	0.89	0.39	0.91
	Ext MCMI	0.23	0.03	14.15	0.10	0.11	0.09	-12.75	0.90	0.90	0.35	0.83
	MI Reg	0.23	0.03	13.53	0.10	0.11	0.11	4.60	0.94	0.94	0.41	0.97
	MI PMM	0.22	0.02	11.47	0.10	0.10	0.11	8.49	0.94	0.94	0.42	0.99
	MI Cart	0.18	-0.02	-8.13	0.10	0.10	0.10	3.54	0.94	0.94	0.39	0.92
	MI RF	0.17	-0.03	-16.32	0.08	0.08	0.11	36.44	0.98	0.99	0.41	0.96
	MI Amelia	0.20	-0.00	-2.13	0.08	0.08	0.09	20.43	0.97	0.97	0.35	0.84
	MI 2Int	0.19	-0.01	-5.70	0.10	0.10	0.11	8.61	0.95	0.96	0.42	1.00
	MI 3Int	0.19	-0.01	-4.17	0.10	0.10	0.11	9.75	0.95	0.95	0.42	0.98
	MI 2IntN	0.18	-0.02	-7.55	0.10	0.10	0.11	11.77	0.96	0.95	0.41	0.98
	MI 3IntN	0.18	-0.02	-8.29	0.09	0.10	0.11	12.04	0.94	0.94	0.41	0.97
B	CC	0.21	0.01	4.49	0.13	0.13	0.12	-10.48	0.91	0.91	0.46	0.93
	Ext	0.21	0.01	5.68	0.14	0.14	0.12	-12.98	0.90	0.90	0.46	0.93
	Ext MCMI	0.22	0.02	11.36	0.12	0.12	0.11	-12.97	0.89	0.90	0.41	0.83
	MI Reg	0.21	0.01	4.76	0.12	0.12	0.12	7.85	0.94	0.95	0.48	0.96
	MI PMM	0.21	0.01	2.60	0.12	0.12	0.13	10.36	0.94	0.94	0.49	0.98
	MI Cart	0.16	-0.04	-17.98	0.10	0.11	0.12	13.63	0.94	0.97	0.44	0.89
	MI RF	0.13	-0.07	-32.57	0.08	0.10	0.12	47.06	0.96	0.99	0.45	0.91
	MI Amelia	0.18	-0.02	-10.02	0.08	0.08	0.10	27.20	0.98	0.98	0.40	0.79
	MI 2Int	0.18	-0.02	-10.27	0.12	0.12	0.13	8.79	0.94	0.94	0.50	1.00
	MI 3Int	0.18	-0.02	-7.68	0.11	0.11	0.13	10.86	0.95	0.95	0.48	0.96
	MI 2IntN	0.17	-0.03	-14.53	0.11	0.11	0.12	14.41	0.95	0.96	0.47	0.95
	MI 3IntN	0.17	-0.03	-13.74	0.11	0.11	0.12	15.25	0.96	0.96	0.47	0.94
C	CC	0.19	-0.01	-5.04	0.13	0.13	0.12	-11.27	0.90	0.91	0.46	0.93
	Ext	0.19	-0.01	-5.47	0.14	0.14	0.12	-13.96	0.90	0.90	0.46	0.93
	Ext MCMI	0.20	-0.00	-1.36	0.12	0.12	0.11	-13.25	0.90	0.90	0.41	0.83
	MI Reg	0.20	0.00	0.02	0.11	0.11	0.13	12.80	0.95	0.95	0.49	0.99
	MI PMM	0.20	-0.00	-1.71	0.11	0.11	0.13	14.65	0.96	0.97	0.49	0.99
	MI Cart	0.16	-0.04	-21.71	0.10	0.11	0.12	17.76	0.95	0.97	0.45	0.90
	MI RF	0.13	-0.07	-35.28	0.08	0.11	0.12	50.02	0.97	0.99	0.46	0.92
	MI Amelia	0.19	-0.01	-6.23	0.08	0.08	0.10	27.92	0.98	0.98	0.40	0.80
	MI 2Int	0.18	-0.02	-9.02	0.12	0.12	0.13	8.59	0.96	0.95	0.49	1.00
	MI 3Int	0.19	-0.01	-6.16	0.11	0.11	0.12	9.70	0.95	0.94	0.47	0.96
	MI 2IntN	0.17	-0.03	-17.26	0.10	0.11	0.12	18.85	0.96	0.97	0.48	0.96
	MI 3IntN	0.17	-0.03	-16.29	0.10	0.11	0.12	16.74	0.96	0.97	0.47	0.95
D	CC	0.21	0.01	5.96	0.12	0.12	0.10	-10.74	0.92	0.92	0.41	0.89
	Ext	0.21	0.01	5.78	0.12	0.12	0.11	-12.09	0.92	0.92	0.41	0.90
	Ext MCMI	0.23	0.03	16.40	0.11	0.11	0.10	-11.76	0.89	0.92	0.37	0.81
	MI Reg	0.24	0.04	18.00	0.11	0.11	0.11	9.23	0.94	0.96	0.44	0.97
	MI PMM	0.23	0.03	15.42	0.10	0.11	0.12	11.51	0.95	0.95	0.45	0.98
	MI Cart	0.19	-0.01	-3.99	0.10	0.10	0.11	8.53	0.95	0.95	0.42	0.92
	MI RF	0.17	-0.03	-16.43	0.08	0.09	0.11	38.25	0.98	0.98	0.43	0.94
	MI Amelia	0.20	0.00	0.51	0.08	0.08	0.10	25.89	0.98	0.98	0.38	0.82
	MI 2Int	0.20	0.00	1.45	0.10	0.10	0.12	14.70	0.95	0.95	0.46	1.00
	MI 3Int	0.20	0.00	2.37	0.10	0.10	0.12	12.55	0.96	0.96	0.45	0.98
	MI 2IntN	0.20	-0.00	-2.12	0.10	0.10	0.12	15.14	0.96	0.96	0.44	0.97
	MI 3IntN	0.19	-0.01	-3.63	0.10	0.10	0.12	17.81	0.96	0.96	0.44	0.97
E	CC	0.21	0.01	5.99	0.15	0.15	0.12	-15.57	0.89	0.89	0.48	0.90
	Ext	0.21	0.01	7.10	0.15	0.15	0.12	-17.99	0.88	0.88	0.48	0.91
	Ext MCMI	0.22	0.02	11.86	0.13	0.13	0.11	-15.46	0.90	0.90	0.43	0.81
	MI Reg	0.23	0.03	15.01	0.12	0.13	0.13	8.28	0.95	0.95	0.51	0.96
	MI PMM	0.23	0.03	14.96	0.12	0.13	0.14	12.56	0.95	0.95	0.53	1.00
	MI Cart	0.18	-0.02	-10.02	0.11	0.11	0.12	15.84	0.95	0.96	0.47	0.89
	MI RF	0.14	-0.06	-29.25	0.09	0.10	0.13	44.49	0.97	0.99	0.48	0.92
	MI Amelia	0.19	-0.01	-4.05	0.09	0.09	0.11	26.60	0.98	0.98	0.42	0.79
	MI 2Int	0.21	0.01	3.61	0.13	0.13	0.13	7.03	0.94	0.95	0.52	0.98
	MI 3Int	0.21	0.01	7.24	0.12	0.13	0.13	6.09	0.95	0.94	0.51	0.96
	MI 2IntN	0.20	-0.00	-1.59	0.11	0.11	0.13	16.93	0.96	0.95	0.51	0.97
	MI 3IntN	0.20	-0.00	-1.48	0.11	0.11	0.13	14.04	0.96	0.95	0.50	0.95
F	CC	0.20	-0.00	-0.97	0.12	0.12	0.11	-9.77	0.92	0.92	0.41	0.89
	Ext	0.20	-0.00	-0.63	0.12	0.12	0.11	-9.35	0.91	0.91	0.42	0.90
	Ext MCMI	0.22	0.02	9.77	0.11	0.11	0.10	-10.68	0.92	0.91	0.37	0.81
	MI Reg	0.23	0.03	13.25	0.10	0.11	0.12	11.60	0.95	0.96	0.45	0.97
	MI PMM	0.22	0.02	12.12	0.11	0.11	0.12	10.60	0.95	0.95	0.45	0.98
	MI Cart	0.19	-0.01	-7.44	0.10	0.10	0.11	13.49	0.96	0.97	0.43	0.93
	MI RF	0.16	-0.04	-29.34	0.08	0.09	0.11	41.26	0.98	0.98	0.43	0.94
	MI Amelia	0.21	0.01	4.55	0.08	0.08	0.10	28.59	0.98	0.98	0.38	0.82
	MI 2Int	0.20	-0.00	-0.37	0.11	0.11	0.12	13.19	0.96	0.96	0.46	1.00
	MI 3Int	0.20	0.00	1.63	0.10	0.10	0.12	12.83	0.96	0.97	0.45	0.97
	MI 2IntN	0.19	-0.01	-6.01	0.10	0.10	0.12	16.17	0.96	0.96	0.45	0.97
	MI 3IntN	0.19	-0.01	-6.12	0.10	0.10	0.12	16.81	0.96	0.96	0.45	0.97
G	CC	0.19	-0.01	-3.24	0.13	0.13	0.12	-10.18	0.91	0.91	0.46	0.93
	Ext	0.19	-0.01	-2.58	0.13	0.13	0.12	-11.67	0.92	0.92	0.46	0.93
	Ext MCMI	0.21	0.01	4.95	0.12	0.12	0.11	-9.70	0.91	0.92	0.41	0.83
	MI Reg	0.20	-0.00	-2.09	0.11	0.11	0.13	11.53	0.95	0.95	0.48	0.97
	MI PMM	0.19	-0.01	-3.69	0.11	0.11	0.13	16.83	0.96	0.96	0.50	1.00
	MI Cart	0.16	-0.04	-22.47	0.10	0.11	0.11	16.12	0.94	0.97	0.44	0.89
	MI RF	0.13	-0.07	-37.21	0.08	0.11	0.12	49.47	0.96	0.99	0.45	0.91
	MI Amelia	0.17	-0.03	-15.04	0.08	0.09	0.10	24.93	0.97	0.98	0.39	0.79
	MI 2Int	0.17	-0.03	-16.62	0.12	0.12	0.13	11.65	0.95	0.96	0.49	0.99
	MI 3Int	0.17	-0.03	-14.35	0.11	0.12	0.13	11.15	0.95	0.95	0.48	0.96
	MI 2IntN	0.16	-0.04	-20.62	0.11	0.11	0.12	15.51	0.95	0.97	0.47	0.95
	MI 3IntN	0.16	-0.04	-20.33	0.10	0.11	0.12	18.87	0.96	0.97	0.47	0.94
H	CC	0.17	-0.03	-13.39	0.13	0.13	0.12	-8.07	0.92	0.93	0.46	0.94
	Ext	0.17	-0.03	-12.95	0.13	0.13	0.12	-9.59	0.90	0.92	0.46	0.94
	Ext MCMI	0.18	-0.02	-9.85	0.11	0.11	0.11	-7.25	0.92	0.93	0.41	0.83
	MI Reg	0.18	-0.02	-9.82	0.11	0.11	0.12	16.98	0.96	0.96	0.48	0.98
	MI PMM	0.18	-0.02	-10.37	0.11	0.11	0.13	18.41	0.96	0.96	0.49	1.00
	MI Cart	0.15	-0.05	-27.07	0.09	0.11	0.11	20.87	0.95	0.97	0.44	0.90
	MI RF	0.12	-0.08	-41.54	0.08	0.11	0.12	53.32	0.95	0.99	0.45	0.92
	MI Amelia	0.17	-0.03	-13.06	0.08	0.08	0.10	34.66	0.98	0.98	0.40	0.81
	MI 2Int	0.16	-0.04	-17.85	0.11	0.12	0.13	11.35	0.94	0.96	0.49	0.99
	MI 3Int	0.17	-0.03	-14.62	0.11	0.11	0.12	13.66	0.95	0.95	0.48	0.97
	MI 2IntN	0.15	-0.05	-25.05	0.10	0.12	0.12	17.39	0.94	0.96	0.47	0.96
	MI 3IntN	0.15	-0.05	-22.62	0.10	0.11	0.12	20.45	0.95	0.97	0.47	0.95
I	CC	0.17	-0.03	-14.39	0.14	0.15	0.12	-13.97	0.89	0.89	0.47	0.90
	Ext	0.17	-0.03	-14.05	0.15	0.15	0.12	-17.30	0.86	0.88	0.48	0.91
	Ext MCMI	0.18	-0.02	-11.31	0.13	0.13	0.11	-15.29	0.88	0.90	0.42	0.81
	MI Reg	0.20	0.00	0.86	0.12	0.12	0.13	11.67	0.95	0.95	0.52	0.99
	MI PMM	0.20	0.00	0.68	0.12	0.12	0.14	13.77	0.96	0.96	0.52	1.00
	MI Cart	0.16	-0.04	-22.22	0.11	0.11	0.12	16.96	0.95	0.97	0.47	0.90
	MI RF	0.12	-0.08	-39.80	0.08	0.12	0.13	50.79	0.96	0.99	0.49	0.92
	MI Amelia	0.19	-0.01	-6.53	0.08	0.08	0.11	29.61	0.97	0.98	0.42	0.79
	MI 2Int	0.19	-0.01	-5.04	0.12	0.12	0.14	10.60	0.96	0.95	0.52	1.00
	MI 3Int	0.20	0.00	0.18	0.12	0.12	0.13	12.33	0.95	0.95	0.51	0.97
	MI 2IntN	0.17	-0.03	-13.46	0.11	0.12	0.13	20.29	0.96	0.96	0.52	0.98
	MI 3IntN	0.18	-0.02	-12.08	0.11	0.11	0.13	16.56	0.95	0.96	0.50	0.96
J	CC	0.16	-0.04	-21.17	0.14	0.14	0.11	-16.51	0.87	0.89	0.45	0.85
	Ext	0.16	-0.04	-21.19	0.14	0.15	0.12	-18.26	0.86	0.88	0.45	0.86
	Ext MCMI	0.16	-0.04	-18.96	0.13	0.13	0.10	-19.37	0.86	0.89	0.40	0.76
	MI Reg	0.19	-0.01	-6.94	0.12	0.12	0.13	6.89	0.94	0.94	0.51	0.97
	MI PMM	0.18	-0.02	-7.87	0.13	0.13	0.14	8.38	0.94	0.94	0.52	1.00
	MI Cart	0.14	-0.06	-29.69	0.11	0.12	0.12	10.89	0.92	0.95	0.45	0.87
	MI RF	0.11	-0.09	-46.20	0.08	0.12	0.12	47.83	0.94	1.00	0.47	0.90
	MI Amelia	0.17	-0.03	-14.36	0.08	0.09	0.11	27.09	0.96	0.97	0.41	0.79
	MI 2Int	0.17	-0.03	-15.72	0.12	0.13	0.14	9.86	0.94	0.95	0.52	1.00
	MI 3Int	0.18	-0.02	-10.03	0.12	0.12	0.13	8.63	0.95	0.96	0.50	0.96
	MI 2IntN	0.16	-0.04	-21.61	0.12	0.12	0.13	12.98	0.94	0.96	0.50	0.96
	MI 3IntN	0.16	-0.04	-20.07	0.11	0.12	0.13	10.08	0.94	0.95	0.48	0.92
T	CC	0.20	0.00	2.19	0.11	0.11	0.10	-10.00	0.92	0.93	0.39	0.92
	Ext	0.20	0.00	1.93	0.11	0.11	0.10	-10.10	0.92	0.92	0.40	0.93
	Ext MCMI	0.22	0.02	10.99	0.10	0.10	0.09	-10.16	0.91	0.92	0.35	0.82
	MI Reg	0.23	0.03	14.74	0.10	0.10	0.11	10.41	0.95	0.96	0.42	0.98
	MI PMM	0.23	0.03	12.75	0.10	0.10	0.11	14.72	0.96	0.97	0.43	1.00
	MI Cart	0.18	-0.02	-8.82	0.09	0.09	0.10	12.88	0.97	0.96	0.40	0.93
	MI RF	0.16	-0.04	-19.48	0.07	0.08	0.10	38.38	0.97	0.99	0.40	0.93
	MI Amelia	0.19	-0.01	-3.74	0.07	0.07	0.10	32.61	0.99	0.99	0.37	0.86
	MI 2Int	0.19	-0.01	-5.84	0.09	0.09	0.11	15.01	0.96	0.96	0.42	0.97
	MI 3Int	0.19	-0.01	-5.50	0.09	0.09	0.11	15.38	0.96	0.96	0.42	0.97
	MI 2IntN	0.19	-0.01	-4.43	0.09	0.09	0.11	14.68	0.97	0.97	0.41	0.96
	MI 3IntN	0.18	-0.02	-7.57	0.09	0.09	0.11	19.32	0.97	0.97	0.42	0.97

Table B.12: Performance of missing data methods with *TMLE* implementation for the 11 assessed missingness mechanisms under the complex 1 scenario for reproduced data

Causal diagram	Missing data method	Estimated mean	Bias	Relative Bias (%)	empSE	RMSE	ModSE	Error in ModSE (%)	Coverage	Bias eliminated coverage	Mean CI length	Proportional CI length
A	CC	0.20	0.00	2.33	0.11	0.11	0.10	-14.08	0.91	0.91	0.37	0.70
	Ext	0.21	0.01	2.91	0.12	0.12	0.10	-15.19	0.90	0.90	0.38	0.71
	Ext MCMI	0.23	0.03	14.48	0.10	0.11	0.09	-15.46	0.88	0.89	0.34	0.65
	MI Reg	0.29	0.09	45.25	0.11	0.14	0.13	24.70	0.92	0.98	0.51	0.97
	MI PMM	0.28	0.08	41.86	0.11	0.14	0.14	25.50	0.93	0.98	0.53	0.99
	MI Cart	0.20	-0.00	-1.85	0.09	0.09	0.12	32.87	0.98	0.98	0.47	0.88
	MI RF	0.18	-0.02	-10.13	0.08	0.08	0.13	59.47	1.00	1.00	0.50	0.95
	MI Amelia	0.22	0.02	12.22	0.08	0.08	0.11	38.09	0.98	0.99	0.41	0.77
	MI 2Int	0.22	0.02	8.91	0.10	0.10	0.14	39.03	0.98	0.98	0.53	1.00
	MI 3Int	0.22	0.02	11.12	0.10	0.10	0.13	35.69	0.98	0.98	0.51	0.96
	MI 2IntN	0.22	0.02	9.87	0.10	0.10	0.13	39.21	0.98	0.98	0.51	0.96
	MI 3IntN	0.22	0.02	8.85	0.09	0.10	0.13	37.78	0.99	0.98	0.50	0.93
B	CC	0.20	0.00	0.46	0.13	0.13	0.11	-14.98	0.89	0.89	0.44	0.75
	Ext	0.20	0.00	2.35	0.14	0.14	0.11	-18.99	0.87	0.87	0.44	0.74
	Ext MCMI	0.22	0.02	12.01	0.12	0.13	0.10	-19.05	0.87	0.88	0.39	0.66
	MI Reg	0.26	0.06	28.15	0.12	0.13	0.15	23.32	0.94	0.96	0.57	0.97
	MI PMM	0.26	0.06	31.19	0.12	0.14	0.16	26.49	0.96	0.97	0.59	1.00
	MI Cart	0.18	-0.02	-11.71	0.10	0.10	0.13	25.76	0.97	0.97	0.49	0.82
	MI RF	0.14	-0.06	-29.33	0.08	0.10	0.14	67.64	0.99	0.99	0.55	0.92
	MI Amelia	0.20	0.00	0.35	0.09	0.09	0.11	34.29	0.98	0.98	0.44	0.75
	MI 2Int	0.21	0.01	3.20	0.12	0.12	0.15	27.76	0.97	0.97	0.58	0.98
	MI 3Int	0.22	0.02	7.60	0.11	0.11	0.15	34.43	0.98	0.98	0.56	0.94
	MI 2IntN	0.19	-0.01	-2.92	0.11	0.11	0.14	34.38	0.98	0.98	0.55	0.93
	MI 3IntN	0.20	-0.00	-1.52	0.10	0.10	0.14	35.45	0.98	0.98	0.53	0.90
C	CC	0.17	-0.03	-13.54	0.13	0.14	0.11	-15.80	0.88	0.89	0.44	0.73
	Ext	0.18	-0.02	-10.97	0.14	0.14	0.11	-18.15	0.88	0.88	0.44	0.73
	Ext MCMI	0.19	-0.01	-4.00	0.12	0.12	0.10	-19.00	0.87	0.87	0.39	0.64
	MI Reg	0.25	0.05	24.35	0.13	0.13	0.15	19.67	0.94	0.96	0.58	0.96
	MI PMM	0.25	0.05	23.43	0.13	0.14	0.16	20.93	0.96	0.97	0.60	1.00
	MI Cart	0.17	-0.03	-15.89	0.11	0.11	0.13	21.19	0.96	0.97	0.49	0.82
	MI RF	0.14	-0.06	-31.79	0.08	0.11	0.14	66.56	0.99	1.00	0.54	0.90
	MI Amelia	0.21	0.01	4.86	0.09	0.09	0.12	34.30	0.98	0.98	0.45	0.74
	MI 2Int	0.21	0.01	2.75	0.12	0.12	0.15	29.55	0.97	0.97	0.59	0.97
	MI 3Int	0.21	0.01	7.30	0.11	0.11	0.14	24.92	0.96	0.96	0.54	0.90
	MI 2IntN	0.18	-0.02	-7.95	0.11	0.11	0.14	32.01	0.98	0.98	0.55	0.91
	MI 3IntN	0.19	-0.01	-4.97	0.11	0.11	0.14	32.95	0.98	0.98	0.54	0.89
D	CC	0.20	-0.00	-0.37	0.11	0.11	0.10	-11.57	0.91	0.91	0.39	0.71
	Ext	0.20	-0.00	-0.66	0.12	0.12	0.10	-12.77	0.90	0.90	0.39	0.71
	Ext MCMI	0.22	0.02	11.55	0.11	0.11	0.09	-13.39	0.90	0.90	0.36	0.64
	MI Reg	0.28	0.08	42.40	0.11	0.14	0.14	27.37	0.93	0.97	0.54	0.97
	MI PMM	0.28	0.08	39.46	0.11	0.14	0.15	30.10	0.95	0.97	0.56	1.00
	MI Cart	0.20	-0.00	-0.07	0.10	0.10	0.13	29.79	0.98	0.98	0.48	0.86
	MI RF	0.18	-0.02	-10.90	0.08	0.09	0.13	60.49	0.99	0.99	0.52	0.93
	MI Amelia	0.22	0.02	12.17	0.08	0.08	0.11	43.31	0.98	0.98	0.43	0.77
	MI 2Int	0.22	0.02	10.52	0.10	0.11	0.14	36.87	0.98	0.98	0.55	0.98
	MI 3Int	0.23	0.03	13.37	0.10	0.10	0.14	41.40	0.98	0.98	0.53	0.96
	MI 2IntN	0.22	0.02	10.37	0.10	0.10	0.14	45.23	0.99	0.99	0.53	0.96
	MI 3IntN	0.22	0.02	8.89	0.10	0.10	0.14	44.16	0.99	0.99	0.53	0.95
E	CC	0.20	-0.00	-2.08	0.14	0.14	0.12	-16.79	0.89	0.89	0.45	0.73
	Ext	0.20	-0.00	-0.34	0.15	0.15	0.12	-21.76	0.87	0.87	0.45	0.73
	Ext MCMI	0.21	0.01	6.69	0.13	0.13	0.10	-18.61	0.88	0.88	0.40	0.65
	MI Reg	0.27	0.07	36.33	0.13	0.15	0.16	20.47	0.93	0.96	0.60	0.98
	MI PMM	0.27	0.07	35.44	0.13	0.15	0.16	22.44	0.94	0.97	0.62	1.00
	MI Cart	0.18	-0.02	-10.71	0.11	0.11	0.13	27.41	0.97	0.97	0.52	0.84
	MI RF	0.14	-0.06	-28.92	0.09	0.11	0.15	64.58	0.99	1.00	0.56	0.91
	MI Amelia	0.21	0.01	3.05	0.09	0.09	0.12	37.42	0.98	0.99	0.47	0.77
	MI 2Int	0.22	0.02	9.31	0.12	0.12	0.16	28.33	0.97	0.97	0.60	0.98
	MI 3Int	0.22	0.02	12.08	0.12	0.12	0.15	27.89	0.97	0.97	0.58	0.94
	MI 2IntN	0.21	0.01	2.86	0.12	0.12	0.15	29.38	0.98	0.97	0.58	0.94
	MI 3IntN	0.21	0.01	5.94	0.11	0.11	0.15	31.21	0.98	0.98	0.56	0.92
F	CC	0.19	-0.01	-4.95	0.12	0.12	0.10	-16.07	0.88	0.89	0.40	0.70
	Ext	0.19	-0.01	-5.23	0.12	0.13	0.10	-17.07	0.89	0.90	0.40	0.71
	Ext MCMI	0.22	0.02	8.81	0.11	0.11	0.09	-17.64	0.88	0.88	0.36	0.64
	MI Reg	0.29	0.09	42.64	0.12	0.15	0.14	21.14	0.93	0.96	0.55	0.97
	MI PMM	0.28	0.08	40.78	0.12	0.14	0.15	24.38	0.94	0.96	0.56	1.00
	MI Cart	0.20	0.00	0.18	0.10	0.10	0.13	26.95	0.98	0.98	0.48	0.86
	MI RF	0.17	-0.03	-13.35	0.09	0.09	0.14	53.57	0.99	1.00	0.52	0.93
	MI Amelia	0.24	0.04	18.65	0.08	0.09	0.11	39.70	0.98	0.98	0.43	0.77
	MI 2Int	0.23	0.03	14.42	0.11	0.11	0.15	34.45	0.97	0.98	0.56	0.99
	MI 3Int	0.24	0.04	18.00	0.11	0.11	0.14	30.39	0.97	0.97	0.53	0.95
	MI 2IntN	0.22	0.02	10.51	0.10	0.11	0.14	31.61	0.98	0.98	0.53	0.93
	MI 3IntN	0.22	0.02	9.56	0.10	0.10	0.14	34.68	0.98	0.97	0.52	0.92
G	CC	0.19	-0.01	-7.49	0.13	0.13	0.11	-12.11	0.91	0.92	0.43	0.74
	Ext	0.19	-0.01	-6.89	0.13	0.13	0.11	-14.88	0.90	0.90	0.43	0.74
	Ext MCMI	0.21	0.01	4.30	0.11	0.11	0.10	-13.51	0.91	0.91	0.38	0.66
	MI Reg	0.24	0.04	22.31	0.12	0.13	0.15	22.28	0.96	0.97	0.57	0.97
	MI PMM	0.25	0.05	23.44	0.12	0.13	0.15	24.40	0.96	0.97	0.58	1.00
	MI Cart	0.16	-0.04	-17.96	0.10	0.11	0.13	27.22	0.97	0.98	0.48	0.83
	MI RF	0.13	-0.07	-33.53	0.08	0.11	0.14	68.23	0.98	0.99	0.54	0.92
	MI Amelia	0.19	-0.01	-5.38	0.08	0.08	0.12	39.15	0.98	0.99	0.44	0.76
	MI 2Int	0.19	-0.01	-4.03	0.12	0.12	0.15	29.07	0.97	0.97	0.58	0.99
	MI 3Int	0.20	0.00	0.29	0.11	0.11	0.14	31.41	0.97	0.97	0.55	0.94
	MI 2IntN	0.18	-0.02	-9.24	0.11	0.11	0.14	35.57	0.98	0.98	0.55	0.94
	MI 3IntN	0.18	-0.02	-8.34	0.10	0.10	0.14	34.96	0.98	0.98	0.53	0.91
H	CC	0.16	-0.04	-18.94	0.13	0.13	0.11	-12.06	0.89	0.92	0.44	0.73
	Ext	0.16	-0.04	-18.10	0.13	0.14	0.11	-15.26	0.88	0.90	0.43	0.73
	Ext MCMI	0.18	-0.02	-10.54	0.12	0.12	0.10	-15.04	0.88	0.90	0.39	0.65
	MI Reg	0.23	0.03	15.73	0.12	0.12	0.15	24.96	0.97	0.97	0.57	0.97
	MI PMM	0.23	0.03	16.82	0.12	0.13	0.16	26.50	0.96	0.98	0.59	1.00
	MI Cart	0.16	-0.04	-21.63	0.10	0.11	0.13	29.31	0.97	0.98	0.49	0.82
	MI RF	0.13	-0.07	-36.56	0.08	0.11	0.14	72.77	0.98	0.99	0.54	0.91
	MI Amelia	0.20	0.00	0.18	0.08	0.08	0.12	38.61	0.99	0.99	0.45	0.75
	MI 2Int	0.19	-0.01	-3.66	0.11	0.12	0.15	33.01	0.98	0.98	0.58	0.98
	MI 3Int	0.21	0.01	2.63	0.11	0.11	0.14	29.14	0.98	0.98	0.54	0.92
	MI 2IntN	0.18	-0.02	-11.02	0.10	0.11	0.14	38.96	0.98	0.98	0.55	0.92
	MI 3IntN	0.18	-0.02	-9.26	0.10	0.10	0.14	35.49	0.98	0.98	0.53	0.90
I	CC	0.17	-0.03	-16.41	0.14	0.14	0.12	-15.52	0.87	0.89	0.45	0.71
	Ext	0.17	-0.03	-14.75	0.14	0.15	0.11	-19.50	0.86	0.88	0.44	0.70
	Ext MCMI	0.18	-0.02	-10.36	0.12	0.13	0.10	-17.88	0.89	0.89	0.40	0.63
	MI Reg	0.25	0.05	27.00	0.13	0.14	0.16	21.06	0.95	0.96	0.61	0.96
	MI PMM	0.25	0.05	25.12	0.13	0.14	0.17	24.57	0.96	0.97	0.63	1.00
	MI Cart	0.17	-0.03	-14.64	0.10	0.11	0.14	33.07	0.98	0.98	0.52	0.83
	MI RF	0.13	-0.07	-36.68	0.08	0.11	0.15	73.94	0.98	1.00	0.56	0.90
	MI Amelia	0.21	0.01	4.07	0.09	0.09	0.12	39.77	0.99	0.99	0.47	0.75
	MI 2Int	0.21	0.01	7.27	0.12	0.12	0.16	30.56	0.97	0.97	0.61	0.97
	MI 3Int	0.22	0.02	12.20	0.12	0.12	0.15	31.69	0.97	0.98	0.58	0.92
	MI 2IntN	0.20	-0.00	-1.22	0.11	0.11	0.15	32.65	0.97	0.97	0.58	0.92
	MI 3IntN	0.20	-0.00	-0.18	0.11	0.11	0.15	36.76	0.98	0.98	0.57	0.91
J	CC	0.15	-0.05	-22.22	0.13	0.14	0.11	-14.29	0.88	0.90	0.43	0.71
	Ext	0.15	-0.05	-22.87	0.13	0.14	0.11	-16.88	0.86	0.89	0.42	0.70
	Ext MCMI	0.16	-0.04	-19.28	0.12	0.12	0.10	-18.43	0.88	0.89	0.38	0.62
	MI Reg	0.23	0.03	16.49	0.13	0.13	0.15	19.89	0.96	0.97	0.59	0.98
	MI PMM	0.24	0.04	17.66	0.13	0.13	0.16	23.15	0.96	0.96	0.60	0.99
	MI Cart	0.16	-0.04	-22.32	0.10	0.11	0.13	32.68	0.97	0.98	0.50	0.83
	MI RF	0.12	-0.08	-40.45	0.08	0.12	0.14	69.36	0.98	0.99	0.54	0.89
	MI Amelia	0.20	-0.00	-2.08	0.09	0.09	0.12	39.63	0.99	0.99	0.46	0.76
	MI 2Int	0.20	0.00	0.48	0.12	0.12	0.16	34.44	0.98	0.98	0.60	1.00
	MI 3Int	0.21	0.01	6.26	0.12	0.12	0.15	30.05	0.97	0.98	0.58	0.96
	MI 2IntN	0.18	-0.02	-8.02	0.11	0.11	0.15	37.05	0.97	0.98	0.57	0.94
	MI 3IntN	0.19	-0.01	-6.00	0.11	0.11	0.14	34.23	0.98	0.98	0.55	0.91
T	CC	0.21	0.01	5.55	0.13	0.13	0.10	-21.19	0.88	0.88	0.39	0.71
	Ext	0.21	0.01	5.55	0.13	0.13	0.10	-21.91	0.87	0.86	0.39	0.72
	Ext MCMI	0.23	0.03	15.93	0.11	0.12	0.09	-21.20	0.85	0.88	0.35	0.64
	MI Reg	0.32	0.12	58.49	0.11	0.16	0.14	23.17	0.89	0.98	0.53	0.97
	MI PMM	0.31	0.11	56.28	0.11	0.16	0.14	27.90	0.90	0.97	0.55	1.00
	MI Cart	0.20	-0.00	-0.79	0.09	0.09	0.13	35.40	0.98	0.98	0.49	0.89
	MI RF	0.18	-0.02	-10.01	0.08	0.08	0.13	63.50	0.99	0.99	0.51	0.94
	MI Amelia	0.24	0.04	20.01	0.08	0.09	0.11	52.14	0.98	0.99	0.44	0.81
	MI 2Int	0.21	0.01	5.68	0.10	0.10	0.14	41.01	0.98	0.98	0.53	0.96
	MI 3Int	0.21	0.01	6.73	0.10	0.10	0.14	39.89	0.98	0.99	0.52	0.96
	MI 2IntN	0.22	0.02	11.82	0.10	0.10	0.14	42.80	0.98	0.99	0.53	0.96
	MI 3IntN	0.22	0.02	7.97	0.10	0.10	0.14	44.67	0.99	0.99	0.53	0.97

Table B.13: Performance of missing data methods with *TMLE* implementation for the 11 assessed missingness mechanisms under the complex 2 scenario for reproduced data

Declaration of Authenticity

The work contained in this thesis is original and has not been previously submitted for examination which has led to the award of a degree.

To the best of my knowledge and belief, this thesis contains no material previously published or written by another person except where due reference is made.