

Comparison of common multiple imputation approaches: An application of logistic regression with an interaction

Research Methods in Medicine & Health Sciences
2024, Vol. 0(0) 1–14
© The Author(s) 2024



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/26320843231224809
journals.sagepub.com/home/rmm



Matthew J Smith¹ , Matteo Quartagno², Aurelien Belot¹, Bernard Rachet¹ and Edmund Njeru Njagi³

Abstract

Background: Multiple imputation is often used to reduce bias and gain efficiency when there is missing data. The most appropriate imputation method depends on the model the analyst is interested in fitting. We consolidate and compare the performance and ease of use for several commonly implemented imputation approaches.

Methods: Using 1000 simulations, each with 10,000 observations, under six data-generating mechanisms (DGM), we investigate the performance of four methods: (i) 'passive imputation', (ii) 'just another variable' (JAV), (iii) 'stratify-impute-append' (SIA), and (iv) 'substantive model compatible fully conditional specification' (SMCFCS). The application of each method is shown in an empirical example using England-based cancer registry data.

Results: SMCFCS and SIA showed the least biased estimate of the coefficients for the fully, and partially, observed variable and the interaction term. SMCFCS and SIA showed good coverage and low relative error for all DGMS. SMCFCS had a large bias when there was a low prevalence of the fully observed variable in the interaction. SIA performed poorly when the fully observed variable in the interaction had a continuous underlying form.

Conclusion: SMCFCS and SIA give consistent estimation and either can be used in most analyses. SMCFCS performed better than SIA when the fully observed variable in the interaction had an underlying continuous form. Researchers should be cautious when using SMCFCS when there is a low prevalence of the fully observed variable in the interaction.

Keywords

Missing data, multiple imputation, interaction, simulation, substantive-model-compatible, fully-conditional-specification, socioeconomic inequalities

Introduction

Missing data is often a problem in medical research because it can cause biased and inefficient inference of parameter estimates when missing data are improperly handled, including when they are ignored (i.e., a complete case analysis), leading to reporting of incorrect conclusions. The validity of inference from incomplete data depends on the mechanism driving missingness; when used properly and assuming the correct missingness mechanism, multiple imputation is known to reduce bias and improve efficiency and is a common approach for handling missing data.^{1,2}

Several approaches exist for performing multiple imputation when modelling a continuous, binary, or time-to-event outcome along with ways of incorporating non-linear and time-varying effects or effect modification.² Various

studies have explored scenarios that involve partially observed outcomes, and/or covariates, and interactions between partially observed covariates.^{3,4} Commonly used

¹Inequalities in Cancer Outcomes Network, Department of Non-Communicable Diseases, London School of Hygiene and Tropical Medicine, London, UK

²MRC Clinical Trials Unit, Institute of Clinical Trials and Methodology, University College London, London, UK

³Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK

Corresponding author:

Matthew J Smith, Inequalities in Cancer Outcomes Network, Department of Non-Communicable Disease, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK.

Email: matt.smith@lshtm.ac.uk

imputation methods are passive imputation, just another variable (JAV), stratify-impute-append (SIA), and substantive model compatible fully conditional specification (SMCFCS). These approaches can be used in differing settings, with SMCFCS the most flexible of the approaches but, to our knowledge, there is no consolidated comparison of the performance of the approaches and their ease of implementation.

To compare the various imputation approaches, we first introduce a motivating example. Medical research often focuses on regressing a binary outcome Y on several covariates, one of which can be fully observed (henceforth referred as Z) or partially observed (X). Suppose one is constructing a logistic regression model, which is often built with an interaction XZ . For example, a model built to estimate the odds of death (Y) amongst a cohort of cancer patients might include an interaction between the cancer diagnostic stage (X) and a patient's comorbidity status (Z). The model assumes the effect of comorbidity on mortality varies at different stages of cancer. In population-based cancer data sets, cancer diagnostic stage (X) is often poorly recorded, leading to studies requiring a suitable approach for multiple imputation. Suppose data were observed on a sample of individuals for a binary outcome (Y), an exposure (X), and an additional covariate (Z). Suppose also that the data in X was not fully observed (i.e., there is some missing data for some individuals). To ascertain the association between Y and X , adjusted for Z , one could perform a logistic regression analysis on only those individuals who have a complete (fully observed) set of X (i.e., complete case analysis). Assuming the missingness mechanism did not involve the outcome, e.g. if the data were *missing completely at random*, these results would not be biased but would be potentially inefficient. If the data were *missing at random* given the outcome, the complete case analysis would be biased, and inefficient, compared to results after performing multiple imputation.

In context of the motivating example, the simplest possible method is to passively impute the component of the interaction separately; this is known as passive imputation. Von Hippel *et al* (2009) showed that passive imputation will, in general, lead to bias and recommended another approach called 'transform then impute' (i.e., JAV), which considers the interaction as an independent variable in the imputation model.³ Seaman *et al* (2014) demonstrated that unless JAV is used in special cases (i.e., for linear regression and where missing data are assumed missing completely at random) it can also lead to biased inference.⁴ Seaman *et al* further showed that JAV will lead to biases in logistic regression models when imputing a partially observed variable that has a non-linear (e.g., quadratic) effect. However, while results are expected to be similar, it is not known how passive or JAV imputation approaches perform for logistic regression models containing an interaction with a partially observed variable (i.e., XZ).⁴ Von Hippel *et al* also proposed

another approach to impute X : 'stratify, then impute' (we refer to this as Stratify-Impute-Append [SIA]). SIA imputes X within strata (i.e., $Z = 1$ and $Z = 0$) of the fully observed variable in the interaction. Using the example before, this is to impute the partially observed cancer diagnostic stage variable within strata of the fully observed binary comorbidity covariate Z . More recently, Bartlett *et al* (2014) developed an approach to perform multiple imputation by fully conditional specification (FCS) where the substantive model is a linear, logistic, or time-to-event model.⁵ This approach accommodates complex terms in the substantive model (e.g., including interaction terms) whereby making the imputation model compatible with the substantive model: thus termed substantive model compatible fully conditional specification (SMCFCS).

In this article, we aim to consolidate and compare the commonly used imputation approaches for different model specifications. In the Methods section we formally describe the various imputation approaches, specify the simulations, and present results of the performance of the methods. Then we describe an empirical example using England-based cancer registry data of patients diagnosed with diffuse large B-cell lymphoma, we illustrate the application of each imputation method for this real-world setting and present the results from each imputation approach. Lastly, we discuss the performance and applicability of each imputation method.

Methods

Data-generating mechanisms

Let Y_i , X_i and \mathbf{Z}_i denote the values of the binary random variables for the outcome, exposure, and covariates, respectively, for individual i ($i = 1, \dots, n$). Assume that $(Y_1, X_1, \mathbf{Z}_1), \dots, (Y_n, X_n, \mathbf{Z}_n)$ are independently and identically distributed. Let $R_i = 1$ if X_i is observed (i.e. if subject i is a complete case), with $R_i = 0$ otherwise.

Let Z_1 represent treatment (1 = treated, 0 = untreated), Z_2 represents age (centred on the mean and rescaled by a factor of 10), Z_3 represents sex (1 = female, 0 = male), Z_4 is a unit increase in deprivation level (higher is more deprived), Z_5 represents comorbidity (1 = comorbidity, 0 = no comorbidity), and X_1 represents the cancer diagnostic stage (1 = late stage, 0 = early stage). These variables were simulated as $Z_1 \sim \text{Binomial}(0.3)$, $Z_2 \sim N(70, 10)$, $Z_3 \sim \text{Binomial}(0.5)$, $Z_4 \sim \text{Beta}(1, 1.2)$, $Z_5 \sim \text{Binomial}(0.3)$, and $X_1 \sim \text{Binomial}(0.4)$. The outcome Y was specified to mimick values estimated from real data and generated according to the following logistic regression model:

$$\begin{aligned} \text{logit}(P(Y_i = 1)) = & -3 + 0.85Z_{1i} + 1.3Z_{2i} + 0.9Z_{3i} + 1.2Z_{4i} \\ & + 0.9Z_{5i} + 1.4X_{1i} + 1.3Z_{5i}X_{1i} \end{aligned} \quad (1)$$

where -3 is the intercept (β_0).

Missingness was imposed on X_1 using a missing at random (MAR) missing data mechanism. The probability of observing X_1 (i.e., $R_i = 1$) was dependent on the variables in the substantive model (including the outcome) with parameters defined as:

$$\begin{aligned} \text{logit}(P(R_i = 1)) = & \alpha_0 + \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + \alpha_3 Z_{3i} \\ & + \alpha_4 Z_{4i} + \alpha_5 Z_{5i} + \alpha_6 Y_i \end{aligned}$$

where each α_k ($k = 1, \dots, 6$) was set to 1. The intercept (α_0) was set to make the probability of observing X_1 equal to 0.8 (i.e., 20% missing data in diagnostic stage).

We define six data-generating mechanisms (DGM) where a factor is varied one-by-one away from a “base-case” data-generating mechanism (Table 1). The base-case DGM assumes the 20% missing data in X is *missing at random*, the fully observed binary variable (Z) has a prevalence of 30%, and their interaction (i.e., XZ) gives an odds ratio of 1.3 (i.e., amongst those with $Z = 1$, the odds of $Y = 1$ is 1.3 times higher amongst those with $X = 1$ compared to those with $X = 0$). Each DGM uses parametric draws from a known model where the true data-generating model is known.

We simulated 10,000 observations (n_{obs}) containing data based on distributions found within real-world settings of cancer registry data on patients with diffuse large B-cell lymphoma in England.⁶⁻⁸ We chose a large enough sample of repetitions ($n_{sims} = 1000$) such that we obtained a small enough Monte Carlo standard error without unfeasible computational time. The formula for the Monte-Carlo confidence interval around the mean estimate is:⁹

$$p \pm 1.96 * \sqrt{p(1-p)/B}$$

If we substitute p with the nominal coverage probability (i.e., 0.95) and B with the number of simulations ($n_{sim} = 1,000$), the estimated coverage (see Section 2.3) should fall between 93.6% and 96.4% (i.e., approximately 19 times out of 20).

The estimands are the exponential of the fixed effect parameter estimates (i.e., odds ratios) of the substantive logistic regression model (equation (1)) for (i) the fully observed variable (Z_5), (ii) the partially observed variable (X_1), and (iii) the interaction term ($Z_5 X_1$).

Imputation approaches

For each simulated data set, and for each data-generating mechanism (DGM), we performed multiple imputation in the following manner: passive imputation, JAV, SIA, and finally SMC-FCS. We first briefly describe each approach.

In passive imputation the imputation model is specified by the distribution of X given Y and Z . Missing values in X are imputed from this distribution and the values for the interaction XZ are then calculated. Passive imputation ensures that the imputed values of XZ conform to the relationship between X and Z (i.e., that $XZ = X \times Z$). On the other hand, JAV imputation does not ensure this relationship holds. JAV imputation ignores the relationship X and Z and imputes XZ as just another variable. This approach imputes missing values assuming that Y , X , Z , and XZ are separate variables. Therefore, imputed values of XZ will not always be consistent with $X \times Z$ (i.e., for an individual with $Z = 1$, X might be imputed as 1, giving $XZ = 1$, but XZ might be imputed as 0).

As an alternative to handle the interaction term, one could use Stratify-Impute-Append (SIA) This approach imputes values of X within levels of the Z . Imputing separately in levels of Z allows associations between X and Y to differ according to the level of Z . The three approaches mentioned thus far either ignore the compatibility between the imputation model and the substantive model (i.e., passive and JAV) or aim to by-pass the interaction term (i.e., SIA). The last approach, substantive model compatible full conditional specification (SMCFCS) ensures X is imputed using an imputation model that is compatible with the

Table 1. Specifications of each data-generating mechanism.

| | Specifications | | | | |
|--|------------------------|-----------------------------|--|--------------------------------|------------------------------------|
| | Missing data mechanism | Fully observed variable (Z) | Prevalence of fully observed variable (%Z) | Proportion of missingness (%X) | Effect of interaction (odds ratio) |
| Data-generating mechanism ^a | | | | | |
| (Base case) 1 | MAR ^b | Binary | 30% | 20% | 1.3 |
| 2 | — | — | — | — | 1.1 |
| 3 | — | — | — | — | 1.7 |
| 4 | MCAR ^b | — | — | — | — |
| 5 | — | Continuous | — | — | — |
| 6 | — | — | 1% | — | — |

^aEach data-generating mechanism is the same as the 'base case' except for the specification in which there is a specified difference.

^bMAR: missing at random, MCAR: missing completely at random.

substantive model. The compatibility is achievable through rejection sampling, which involves drawing values from a proposal density (i.e., a function of the missing values in X given the other variables) and accepting the draw if it satisfies certain conditions based on a ratio of a target density to the proposal density. The target density incorporates the parameters from the substantive model, allowing for compatibility with the imputation model.

For DGM 1-4, and DGM six (i.e., not including DGM 5), firstly, missing values of X were passively imputed and values for the interaction (XZ) were calculated from the combination of the imputed X values and the observed Z values. Secondly, we imputed missing XY values using the 'just another variable' (JAV) approach, as proposed by von Hippel,³ Thirdly, we performed Stratify-Impute-Append (SIA) imputation by initially stratifying the data by the groups of the fully observed variable (Z) within the interaction (XZ). We then imputed missing values in X , separately, for each stratified data set, and calculated the values for interaction (XZ). The stratified data sets were then appended for calculation of the estimands. Fourthly, we used SMC-FCS to impute missing values in X and XZ .

For DGM 5, the fully observed variable (Z) in the interaction (XZ) was of a continuous form, and passive imputation, JAV, and SMC-FCS were performed as in DGM 1-3 (and DGM 6). However, for SIA, the continuous form of Z was categorised using quintiles (i.e., creating five groups of equal sizes). SIA imputation for X was carried out in 5 separate data sets (i.e., one data set for each of the five groups), which were then combined. The odds ratios (i.e., estimands) were calculated assuming a linear distribution (i.e., continuous form) across the levels of the categorical variable Z .

As a comparison, we also include a "null scenario" where the substantive model does not include an interaction term (i.e., where the effect of the interaction has an odds ratio of 1).

For each imputation method we impute 10 data sets with 10 iterations between imputations. Estimates and confidence intervals (CI) were calculated using Rubin's rules.^{10,1} We used R software for all analyses. The Passive, JAV, and SIA imputation methods were performed using the *mice*¹¹ package and SMC-FCS was performed using the *smcfcs*⁵ package.

Performance measures

We assess the performance of each of the four imputation methods by comparing:

- (i) relative bias (i.e., the relative difference between the mean estimated coefficient, $E[\hat{\theta}]$, and the true value of the coefficient, θ)

$$\text{Relative Bias} = \frac{E[\hat{\theta}] - \theta}{\theta} = \frac{\frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} \hat{\theta}_i - \theta}{\theta}$$

- (ii) coverage (i.e., proportion of 95% confidence intervals that include the true coefficient)

$$\text{Coverage} = \Pr(\hat{\theta}_{\text{low}} \leq \theta \leq \hat{\theta}_{\text{upp}}) = \frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} 1(\hat{\theta}_{\text{low},i} \leq \theta \leq \hat{\theta}_{\text{upp},i})$$

- (iii) relative error is a comparison between the average model-based standard error (ModSE) and the empirical standard error (EmpSE),¹² and expresses how closely the model-based standard error approximates the empirical standard error (EmpSE):

$$\widehat{\text{ModSE}} = \sqrt{\frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} \widehat{\text{Var}}(\hat{\theta}_i)}$$

$$\widehat{\text{EmpSE}} = \sqrt{\frac{1}{n_{\text{sim}} - 1} \sum_{i=1}^{n_{\text{sim}}} (\hat{\theta}_i - \bar{\theta})^2}$$

$$\text{Relative error} = 100 \left(\frac{\widehat{\text{ModSE}}}{\widehat{\text{EmpSE}}} - 1 \right)$$

Simulation results

The results of the performance measures for each imputation approach are shown in the [Figure 1](#) (and [Appendix Table A1](#)).

In the null scenario (i.e., no interaction), there was negligible bias, and optimal coverage, of the fully and partially observed variables for all imputation methods ([Appendix Table A1](#)). The relative error was similar across all four imputation methods for the fully observed variable but SMC-FCS had the lowest relative error for the partially observed variable.

Bias (and relative bias). [Figure 1](#), row 1, shows the results of the bias for the coefficient of the fully observed variable for each imputation method. SIA and SMC-FCS showed negligible bias for DGMs 1-5, but were the most biased for DGM 6. JAV consistently overestimated the effect of the fully observed variable in DGMs 1-4 but severely underestimated the effect in DGM 6.

[Figure 1](#), row 2, shows bias for the coefficient of the partially observed variable for each imputation method. For all four imputation approaches, there was negligible bias, except for SIA, which severely underestimated the effect when the partially observed variable had an underlying

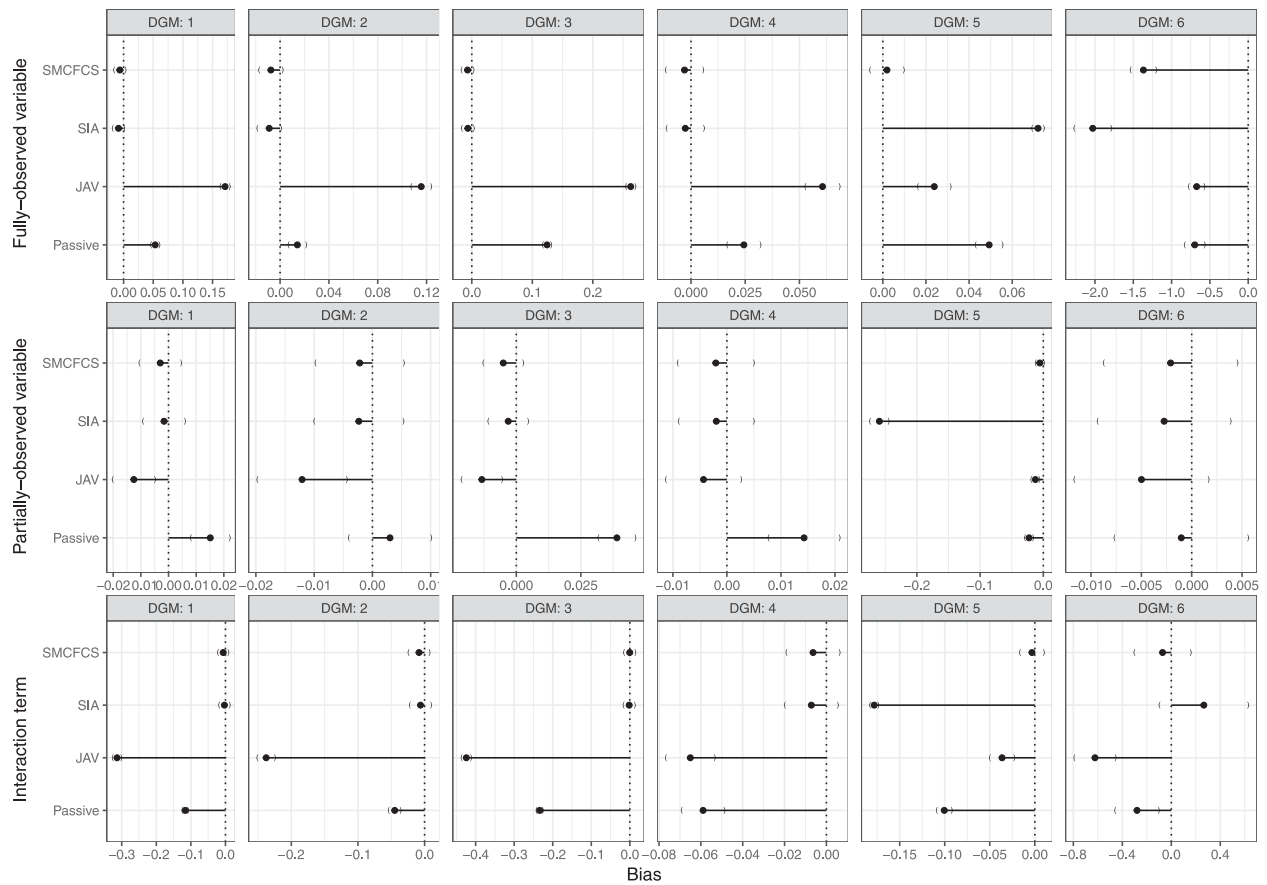


Figure 1. Bias of each imputation method for each variable of interest across simulations ($n = 1000$) by data-generating mechanism.

continuous form (i.e., DGM 5). Passive imputation, was slightly biased in DGMs 1, 3, 4, and 5. JAV showed an underestimate of the partially observed variable in DGMs 1, 2, 3, and 4.

Figure 1, row 3, shows bias for the coefficient of the interaction for each imputation method, Appendix Table A1 shows the precise values. For DGMs 1-4, SIA and SMCFCS showed negligible bias, JAV was severely and the most biased, and passive imputation was biased for DGMs 1-4. In DGM 5 (continuous form of fully observed variable), SMCFCS showed negligible bias, SIA was the least accurate (most biased) followed by passive imputation, then JAV. In DGM 6, JAV had the largest bias, followed by passive imputation and SIA, and SMCFCS was the least biased.

Coverage. Figure 2 shows the results of the coverage of 95% confidence intervals for the interaction variable for each imputation method. Specific values are shown in Appendix Table A1. SMCFCS consistently showed an optimal coverage for DGMs 1-5, and was only affected by overcoverage for DGM 6. SIA had similar results, but for DGM5, where it had severe undercoverage (35.6%). Coverage for the other

two methods was instead inferior, with JAV undercovering in all scenarios but DGM six and passive imputation showing overcoverage for DGM two and 6, and undercoverage for DGM3.

For the partially observed variable, all of the imputation approaches, except SIA in DGM 5, showed a similar coverage in the optimal range (93.6%–96.4%) (Figure 3). SIA had severe undercoverage (75.3%) when the fully observed variable had an underlying continuous form (i.e., DGM5).

For the fully observed variable (Figure 4), results were broadly in line with those for the interaction parameter.

Relative error. For the fully observed variable (Figure 5, row 1), SMCFCS and SIA showed negligible relative error for DGMs 1-4, a larger relative error for DGM 5, and a very large relative error for DGM 6. JAV showed a large relative error for DGMs 1-3, a slightly larger ModSE for DGMs four and 5, and a very large relative error for DGM 6. Passive imputation showed a large relative error for all DGMs.

For the partially observed variable (Figure 5, row 2), SMCFCS, SIA and JAV showed a relative error that was within the expected range (i.e., within MCSE confidence

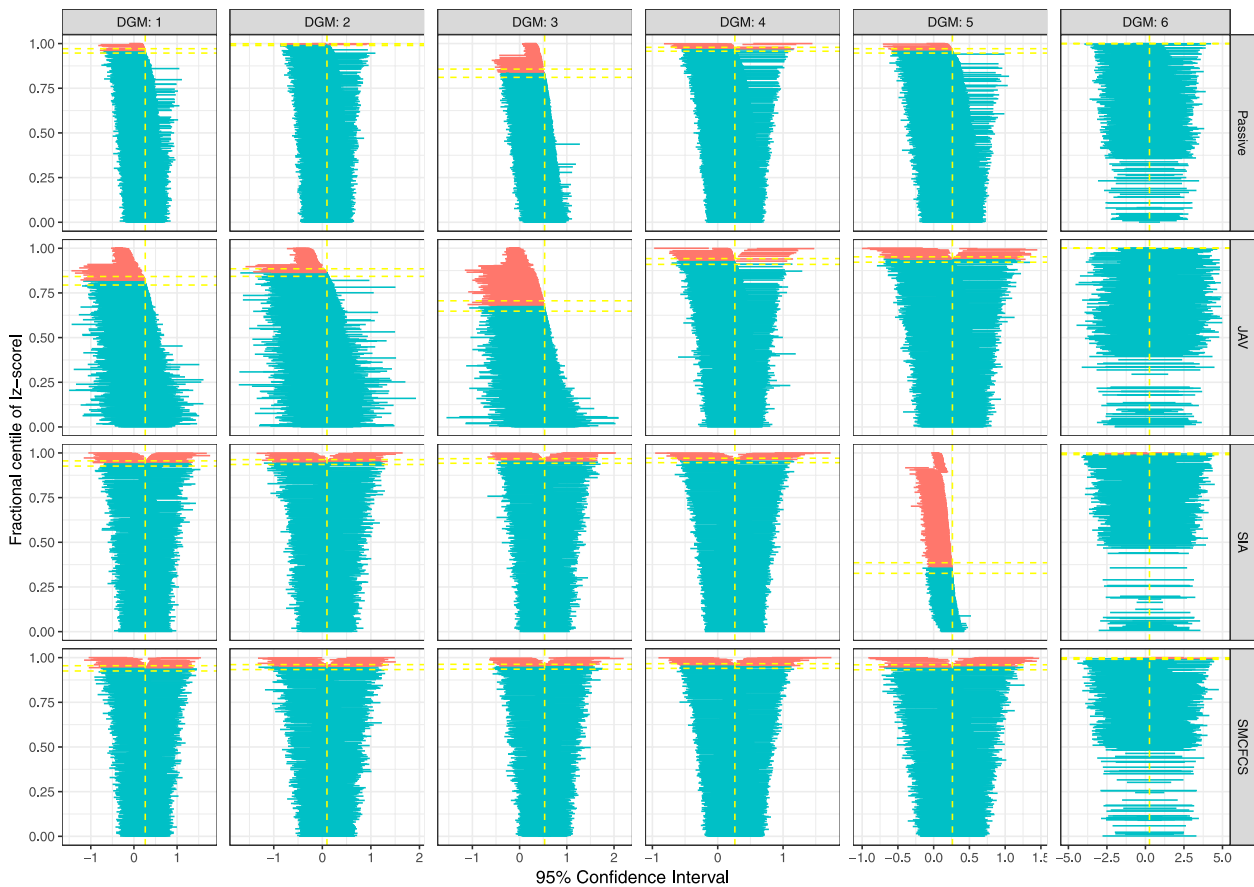


Figure 2. Coverage of 95% confidence intervals for the interaction term across simulations ($n = 1000$) by data-generating mechanism and for each imputation method.

intervals) for all DGMs. Passive imputation showed a slightly larger relative error than expected (i.e., outside the MCSE confidence intervals) for DGMs 1, 2, and 4, but was within MCSE confidence intervals for DGMs 3, 5, and 6.

For the interaction term (Figure 5, row 3), SMCFCOS and SIA showed an optimal relative error (i.e., within the confidence intervals of the MCSE) for DGMs 1-5, but had a very large relative error for DGM 6 (low prevalence of the fully observed variable). JAV's average model-based standard error overestimated the empirical standard error (i.e., large relative error) within DGMs 1-3. JAV showed an optimal relative error for DGM four and DGM 5, and severely overestimated the relative error in DGM 6. Passive imputation showed the largest overestimate of the relative error for all DGMs.

Cancer registry data example

Description of the data

The National Cancer Registry and Analysis Service (NCRAS), run by Public Health England (PHE), is

responsible for cancer registration in England and supports cancer epidemiology, public health, service monitoring and research. Information on patients diagnosed with cancer is essential for assessing the public health system's ability to care for these patients. In England, 17,345 patients were diagnosed with diffuse large B-cell lymphoma between 1st January 2014 and 31st December 2017. Information was available on the patient's age at diagnosis, sex, socioeconomic status, comorbidity status, and cancer stage at diagnosis.

A patient's comorbidity status is based on the Charlson comorbidity index¹³ and was defined as "the existence of disorders, in addition to a primary disease of interest, which are causally unrelated to the primary disease".^{14,15} Comorbidities were coded within Hospital Episode Statistics¹⁶ data according to the International Classification of Diseases, 10th revision.¹⁷ Patients with any previous malignancy were removed. For each patient, we defined a time window of 6–24 months prior to cancer diagnosis for a comorbidity to be recorded. A patient's comorbidity status was determined using an algorithm developed by Maringe et al.¹⁸

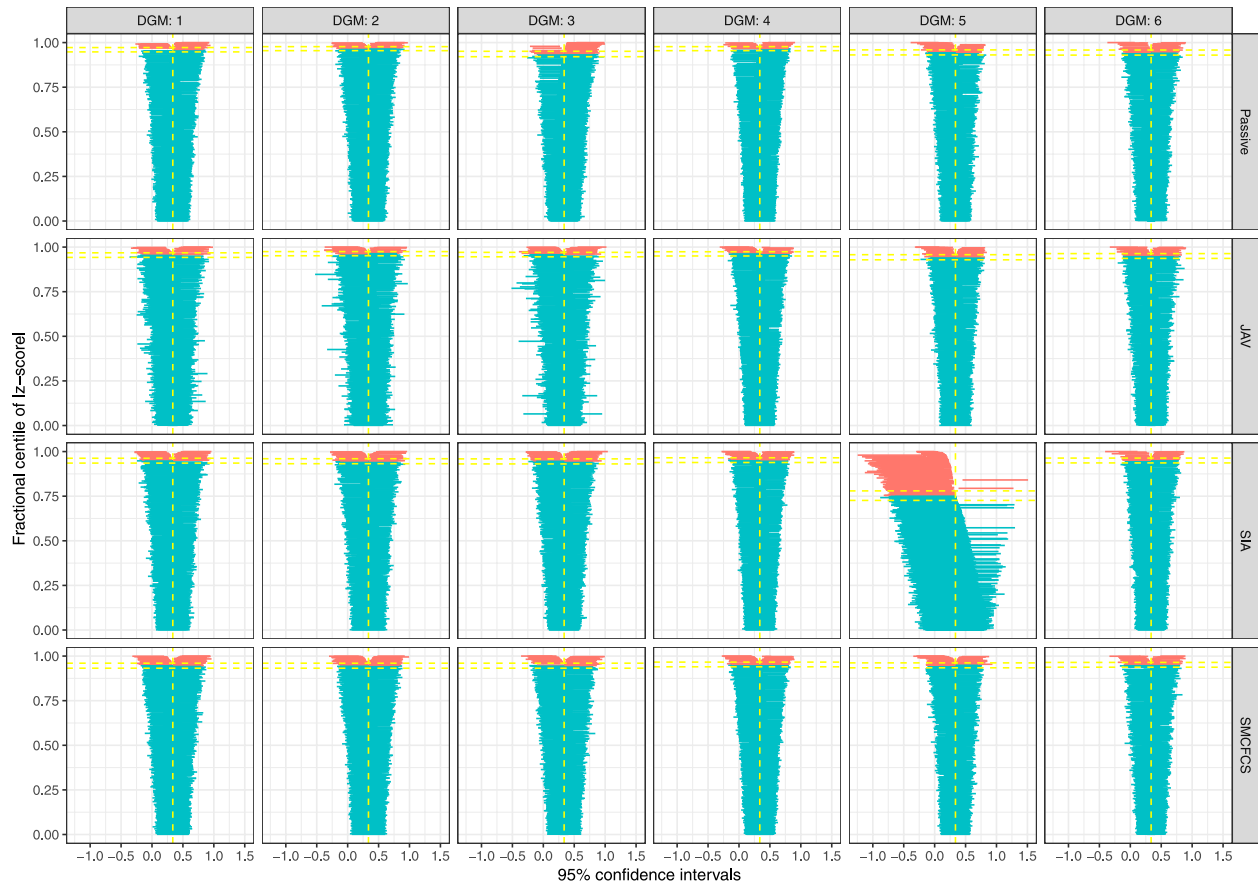


Figure 3. Coverage of 95% confidence intervals for the partially observed variable across simulations ($n = 1000$) by data-generating mechanism and for each imputation method.

It is known that the presence of comorbidity symptoms can delay or hasten the cancer diagnostic stage.¹⁹ For example, comorbidity symptoms similar to the cancer might hide the cancer symptoms, whereas dissimilar symptoms might highlight the differing diagnoses (particularly if the patient is frequently attending routine health appointments for their underlying comorbid condition). For patients with DLBCL, it has been shown that comorbidities are associated with diagnostic delay.^{6,8,20} It would be natural to assume that a statistical model included an interaction between comorbidity and cancer diagnostic stage when investigating the probability of short-term mortality.

We aimed to estimate the parameters of a logistic regression model for the odds of death within 90 days since cancer diagnosis amongst patients diagnosed with diffuse large B-cell lymphoma. Data was available on patient's age (continuous, range: 15.1 - 99.0), sex (binary, Male = 0, Female = 1), deprivation (continuous, range: -1.9 - 6.7), comorbidity (None = 0, at least one = 1), and diagnostic stage (early [stages I/II] = 0, late [stages III/IV] = 1). Information on treatment was not available. Patient characteristics, and the crude odds ratio (OR) of death within 90 days since diagnosis is shown in Table 2.

For the adjusted analysis, the substantive model adjusted for age, sex, deprivation, comorbidity, diagnostic stage, and an interaction between comorbidity and diagnostic stage. Multiple imputation was performed using the four imputation methods (i.e., passive imputation, JAV, SIA, and SMCFCFS). We used 10 imputation with 10 iterations and combined results using Rubin's rules. We tabulate the results of each imputation method in the results section along with the results from a complete case analysis (i.e., ignoring missing data).

Results

Table 3 shows the results of the analysis on the probability of death within 90 days since cancer diagnosis. Under complete case analysis, the interaction term (Odds Ratio [OR] 0.90, 95% confidence interval [CI] 0.71 - 1.14) indicated a possible protective effect on the odds of death within 90 days amongst those with comorbidity and late stage at diagnosis but significant uncertainty remained. Amongst those with no comorbidities, the odds of death within 90 days was 2.51 (95% CI: 2.04 - 3.07) times higher

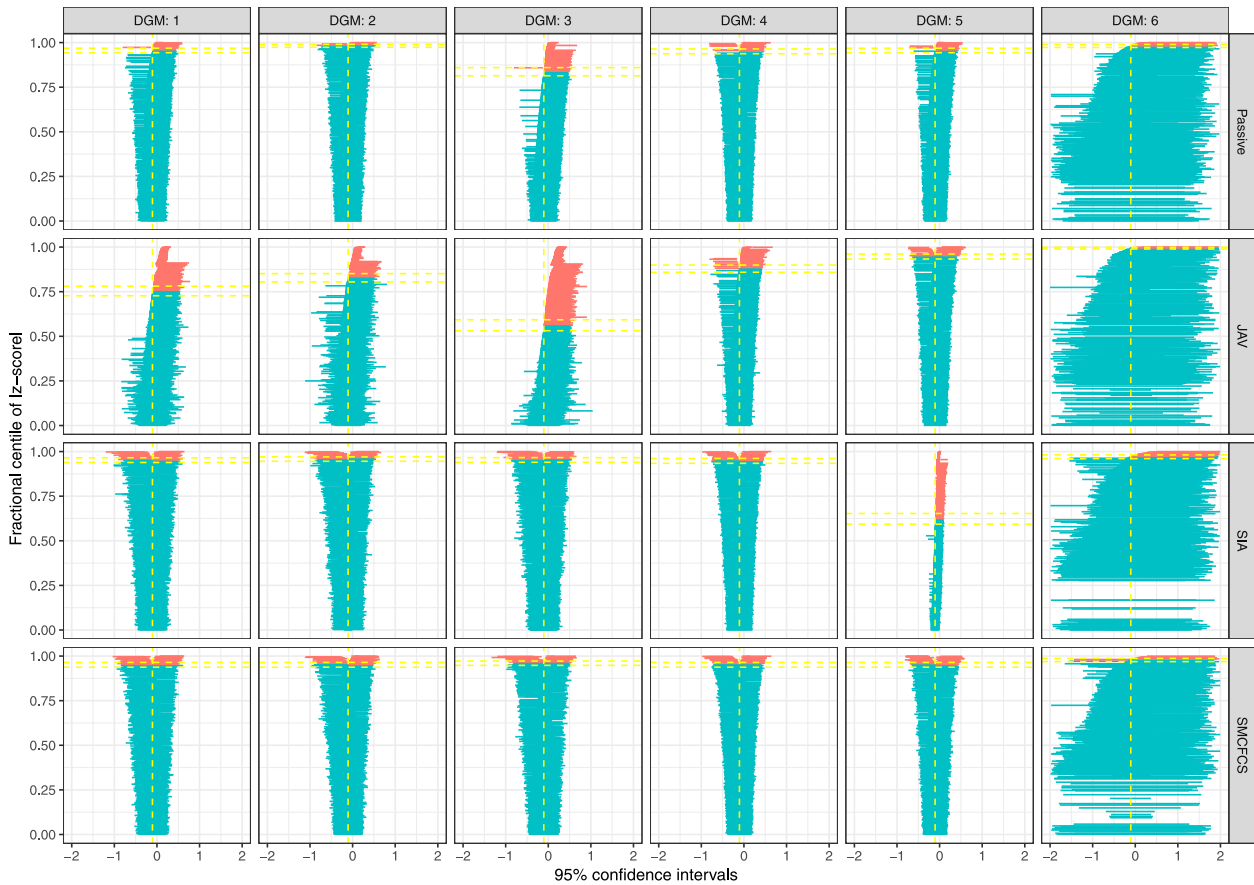


Figure 4. Coverage of 95% confidence intervals for the fully observed variable across simulations ($n = 1000$) by data-generating mechanism and for each imputation method.

for those with a late stage at diagnosis compared to those with an early stage. Amongst those with early diagnostic stage, the presence of comorbidity increased the odds of death by 1.54 (95% CI 1.25 - 1.90) compared to those without comorbidity. For the other variables in the model, the odds of death within 90 days was associated with age (OR 1.93, 95% CI 1.84 - 2.02), sex (OR 0.90, 95% CI 0.82 - 0.99), and deprivation (OR 1.12, 95% CI 1.09 - 1.16), adjusting for the other covariates.

The data from this example compares most closely with DGM two from the simulation study. The missing data in stage is assumed to be missing at random, the fully observed variable is of binary form, the prevalence of the fully observed variable is not small (i.e., 60.7%), and there is a *small* effect of the interaction between comorbidity and diagnostic stage on the probability of death.

For the interaction term, SIA and SMCFCFS showed very similar coefficients and confidence intervals, differing only by at most 2 decimal places. Passive imputation showed a slightly weaker effect of the interaction. JAV showed a strong protective effect (OR 0.73, 95% CI 0.47 - 1.08) of the interaction term. The coefficient for the partially observed

variable (diagnostic stage) was similar between SMCFCFS and SIA, which differed from a similar result between JAV and passive imputation. The coefficient, and confidence intervals, for the fully observed variable (comorbidity status) was again similar between SIA and SMCFCFS, but also similar to passive imputation. JAV showed a much larger harmful effect of comorbidity (OR 1.77, 95% CI 1.28 - 2.44) compared to the other imputation approaches and the complete case analysis.

There was negligible difference between the imputation approaches for the odds of death, and the confidence intervals (CI), for age, sex and deprivation and the substantial conclusions remained the same irrespective of the methods used.

Discussion

We assessed the performance of four imputation approaches when handling missing data before fitting a logistic regression model with an interaction that contains a partially observed variable. We assessed their performance on the coefficients for (i) the interaction term, (ii) the partially

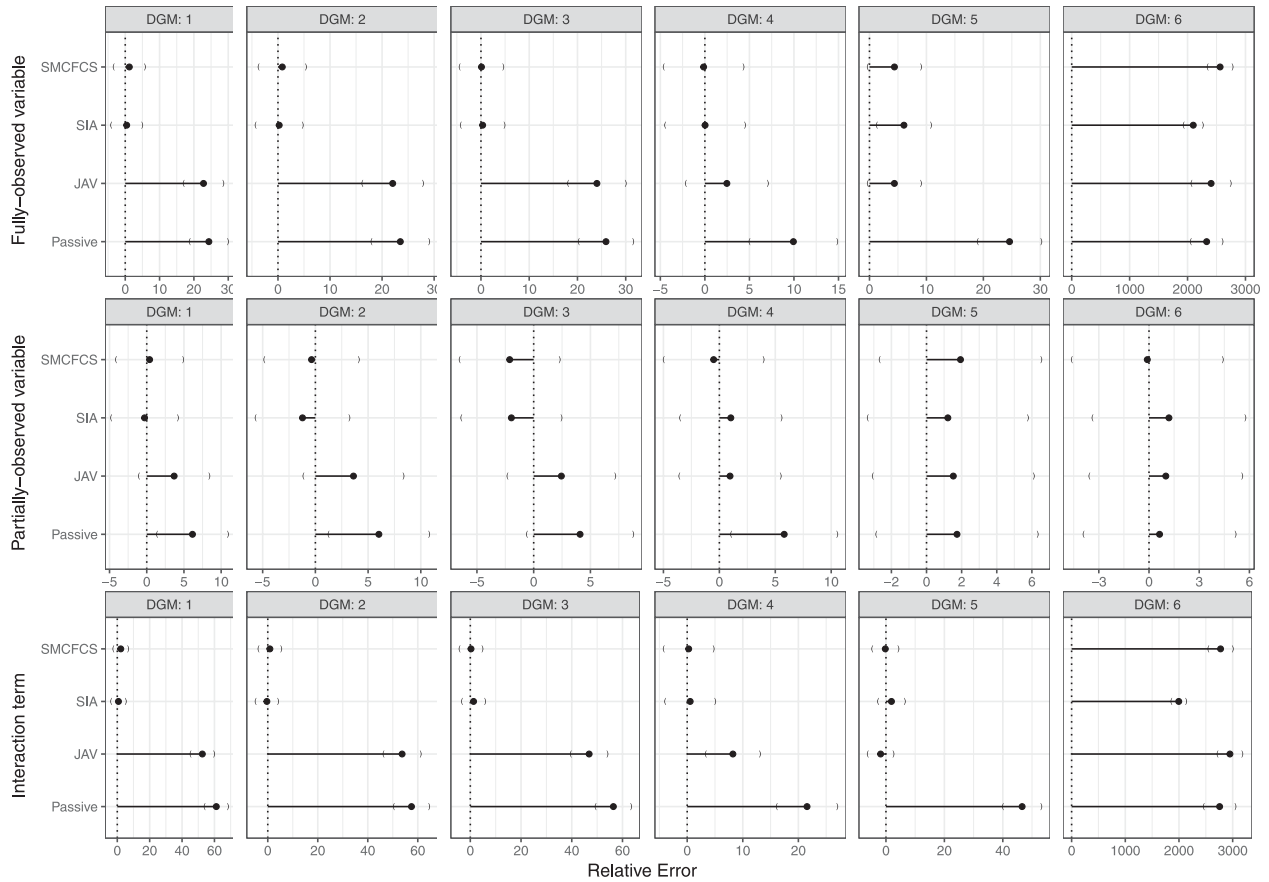


Figure 5. Relative error for the variable of interest across simulations ($n = 1000$) by data-generating mechanism and for each imputation method.

Table 2. Patient characteristics and crude odds ratios of death within 90 days after diagnosis of diffuse large B-cell lymphoma amongst 17,345 patients diagnosed in England between 2014 and 2017.

| | Alive N (%) | Dead N (%) | Crude OR | 95% CI | p -value |
|-------------------------|---------------|--------------|------------|-------------|------------|
| Age^a | | | | | |
| Years (SD) | 67.4 (14.5) | 76.6 (11.3) | 1.83 | 1.78 – 1.91 | <0.001 |
| Sex | | | | | |
| Male | 8025 (55.5) | 1578 (54.8) | Ref | | |
| Female | 6439 (44.5) | 1303 (45.2) | 1.03 | 0.95 – 1.12 | 0.484 |
| Deprivation | | | | | |
| Years (SD) | –0.02 (1.44) | 0.12 (1.50) | 1.07 | 1.04 – 1.10 | <0.001 |
| None | 6009 (41.5) | 808 (28.1) | Ref | | |
| Comorbidity | | | | | |
| At least one | 8455 (58.5) | 2073 (72.0) | 1.82 | 1.67 – 1.99 | <0.001 |
| Diagnostic stage | | | | | |
| Early | 5193 (39.4) | 509 (22.5) | Ref | | |
| Late | 7986 (60.6) | 1751 (77.5) | 2.24 | 2.01 – 2.48 | <0.001 |
| Missing ^b | 1285 (8.9%) | 621 (21.6%) | <i>n/a</i> | <i>n/a</i> | <i>n/a</i> |

SD: standard deviation, OR: odds ratio, 95% CI: confidence interval.

^aAge is centred on its mean (68.9 years) and rescaled by a factor of 10.

^bMissing proportions are calculated separately from observed data.

Table 3. Odds of mortality within 90 days from diagnosis of patients diagnosed with diffuse large B-cell lymphoma in England between 2014 and 2017. JAV: Just-Another Variable. SIA: Stratify-Impute-Append. SMCFCS: substantive model compatible fully conditional specification.

| | Complete case analysis | After multiple imputation | | | | | | | | | |
|--------------------------|------------------------|---------------------------|--------|--------------|--------|--------------|--------|--------------|--------|--------------|--------|
| | | Passive | | | | JAV | | SIA | | SMCFCS | |
| | | OR | 95% CI | OR | 95% CI | OR | 95% CI | OR | 95% CI | OR | 95% CI |
| Age ^a | 1.93 | (1.84, 2.02) | 1.88 | (1.80, 1.96) | 1.87 | (1.79, 1.94) | 1.88 | (1.80, 1.96) | 1.88 | (1.80, 1.96) | |
| Sex | | | | | | | | | | | |
| Male | Ref | — | Ref | — | Ref | — | Ref | — | Ref | — | |
| Female | 0.90 | (0.82, 0.99) | 0.92 | (0.85, 1.00) | 0.92 | (0.85, 1.00) | 0.92 | (0.85, 1.00) | 0.92 | (0.85, 1.00) | |
| Deprivation | 1.12 | (1.09, 1.16) | 1.12 | (1.09, 1.15) | 1.12 | (1.09, 1.15) | 1.12 | (1.09, 1.15) | 1.12 | (1.09, 1.15) | |
| Comorbidity | | | | | | | | | | | |
| None | Ref | — | Ref | — | Ref | — | Ref | — | Ref | — | |
| At least one | 1.54 | (1.25, 1.90) | 1.43 | (1.17, 1.75) | 1.77 | (1.28, 2.44) | 1.45 | (1.18, 1.80) | 1.46 | (1.18, 1.80) | |
| Diagnostic stage | | | | | | | | | | | |
| Early | Ref | — | Ref | — | Ref | — | Ref | — | Ref | — | |
| Late | 2.51 | (2.04, 3.07) | 2.42 | (1.97, 2.97) | 2.43 | (1.85, 3.19) | 2.48 | (2.01, 3.07) | 2.49 | (2.03, 3.05) | |
| Interaction ^b | 0.90 | (0.71, 1.14) | 0.93 | (0.73, 1.18) | 0.72 | (0.47, 1.08) | 0.91 | (0.71, 1.17) | 0.91 | (0.71, 1.16) | |

^aIncrease in odds for each 10-years increase in age.

^bInteraction between stage and comorbidity.

observed variable, and (iii) the fully observed variable in the interaction. SMCFCS imputation consistently provided the least biased estimate of the three coefficients, except when there was a low prevalence of the fully observed variable in which case Just Another Variable (JAV) and passive imputation approaches showed the least bias. The Stratify-Impute-Append (SIA) approach also provided the least biased estimate for the three coefficients of interest, except when the fully observed variable had an underlying continuous form.

SMCFCS showed good or optimal coverage for all scenarios. SIA performed similarly to SMCFCS, except when the fully observed variable had an underlying continuous form. JAV consistently showed large undercoverage for the coefficients of the fully observed variable and the interaction. This is most likely driven by the previously noted large bias, which leads to the severe type II error.¹² For the relative error, SMCFCS consistently provided a close approximation between the average model-based standard error and the empirical standard error, except when there was a low prevalence of the fully observed variable. In this case, SIA approach had the smallest relative error of the four imputation approaches but this was still very large.

Careful consideration should be given to which parameter is of interest to the study when using JAV imputation. Although all four imputation approaches have similar bias when estimating the coefficient of the partially observed variable, JAV imputation will not only be biased for the interaction term but will also introduce bias into the

coefficient for the fully observed variable that is included in the interaction. If the coefficient of the partially observed variable is the primary parameter of interest and the interaction term (and the coefficient of the fully observed variable) is a nuisance parameter, then any of the four methods are appropriate. However, if all three parameters are of interest, as is most often the case, SIA and SMCFCS should be used because they will give the least biased estimates of the slope parameters for X , Z , and XZ .

Seaman *et al* (2012) had previously showed that JAV imputation should not be used for logistic regression models with a *quadratic term*.⁴ Our results are consistent with their conclusions and extend the implications to the case of logistic regression models that include an interaction term. Passive imputation performed better than JAV in our settings, except when the fully observed variable in the interaction had an underlying continuous form, but we have shown that even in the simplest of specifications for a logistic regression model with interactions, both passive or JAV imputation should best be avoided. The only exception was that JAV performed almost as well as SMCFCS when the fully observed variable in the interaction has an underlying continuous form. However, this is a rare and specific example, and other imputation approaches may, more often, perform better.

SMCFCS is possibly the best option for most analyses but SIA is also a suitable alternative, particularly in situations where the model being fit is not available in the SMCFCS package (e.g., excess hazard models in survival analysis). SIA allows one to remove the complexity of the

interaction term and impute the missing data within levels of the categorical variable. SIA can be used when the fully observed variable is a categorical variable but only when there is a large enough number of observations for each level (i.e., avoiding perfect prediction). However, SIA is difficult to use if the fully observed variable in the interaction has a continuous underlying form and one would need to carefully consider how to categorise the continuous variable so that it appropriately captures the patterns observed in the variable; an approach could incorporate machine learning techniques such as classification and regression trees. Moreover, SIA cannot be used when both variables in the interaction contain missing values but SMCFCFS can be used (passive and JAV can also be used but might provide biased results). We focused on multiple imputation because it is the most widely used approach for dealing with missing data, other methods such as inverse probability weighting (IPW) were not considered.^{21,22}

We considered the simple scenario where the only complex covariate was an interaction. Seaman *et al* (2012) investigated the performance of imputation approaches for logistic regression that included a quadratic term as the only complex covariate.⁴ They found that, under MAR assumption, passive imputation produced biased estimates of the quadratic term and predictive mean matching (PMM) was unbiased with correct coverage. Since the SIA approach performs imputation within strata of the fully observed variable (i.e., removing the interaction and imputing only the partially observed covariate), it is possible that SIA can be used for more complex substantive models. For example, a logistic regression model with both a quadratic term (X^2) and an interaction (XZ) could be handled by using predictive mean matching within strata of the fully observed variable of the interaction. The logistic regression of the form $Y = \beta_0 + \beta_1 Z + \beta_2 C + \beta_3 C^2 + \beta_4 X + \beta_5 XZ$, where C is a continuous variable with a quadratic effect on Y , then X and C could be imputed using PMM within strata of Z , yielding unbiased estimates for a complex logistic regression model. This hybrid approach would combine PMM and SIA, which we refer to as "stratified predictive mean matching". The performance of this hybrid approach, and in comparison to SMCFCFS, requires further investigation.

SMCFCFS performed poorly for the bias and coverage of the coefficient of the fully observed variable (i.e., Z_5) in the interaction when there was a low prevalence (i.e., 1%) of this fully observed variable. In this scenario (i.e., DGM 6), the average number of observations for which $Z_5 = 1$ was 100, and the average number of observations for which $Z_5 = 0$ was 9900, in the simulated study. The low prevalence of this variable could lead to perfect prediction for the interaction term if there are low numbers for those with both $Z_5 = 1$ and $X_1 = 1$. Since the prevalence of $X_1 = 1$ was simulated to be 40%, the approximate number of observations for $Z_5 = 1$ and $X_1 = 1$ in this scenario would be 40. SMCFCFS could

be adapted in this scenario with a data augmentation scheme where a few additional carefully crafted observations are added before the imputation model is fitted.²³ Another approach could be to use Firth's bias correction approach, which modifies the maximum likelihood procedure by removing its first order finite sample bias.^{24,25} We also simulated 100,000 observations to specifically assess the performance of SMCFCFS in a larger sample, we found SMCFCFS performed better with larger samples (results not shown). Further studies are required to assess the performance of these adaptations in various scenarios, particularly for low prevalence of the fully, or partially, observed variables.

We compared the results of the four imputation methods using a real-world cancer data set on patients diagnosed with diffuse large B-cell lymphoma in England. As expected, SIA and SMCFCFS showed very similar estimates for the effect (and confidence intervals) of diagnostic stage on mortality with 90 days since diagnosis. Passive imputation produced similar results to the complete case analysis, SIA, and SMCFCFS approaches; however, as shown in the simulations, passive imputation is expected to be biased for the interaction term and the coefficient of the fully observed variable. In the simulations, JAV imputation showed markedly different (and poorer) results for the interaction term and the coefficient of the fully observed variable, which was also reflected in the cancer data analysis and would explain the large bias (away from the null) for these two parameters.

Stage at diagnosis is often considered a categorical variable (i.e., stages I, II, III, and IV). We used a binary form for diagnostic stage (i.e., early vs late stage) for two reasons. Firstly, treatment, such as combination immunochemotherapy, for patients with diffuse large B-cell lymphoma is allocated based on an international prognostic index (IPI).²⁶ One criteria of the DLBCL-IPI is whether the cancer is late stage (i.e., stage III or IV) or early stage (i.e., stage I or II). Secondly, for simplicity of the simulation study, a categorical form for diagnostic stage would be handled similarly to a binary variable (i.e., impute within levels of the fully observed variable: comorbidity) and we anticipate this would not change the conclusions of our results.

We assumed that missing data in diagnostic stage was *missing at random*. In real-world settings, over the past 5 years, the proportion of missing diagnostic stage has reduced. The proportion of patients with a late diagnostic stage has either plateaued or increased at the same rate as the decrease in missing diagnostic stage, possibly indicating the two are related. Thus, the probability of observing cancer stage at diagnosis could be associated with the actual value of the cancer stage. This could potentially be true even after conditioning on all the observed covariates. Thus, the missing cancer diagnostic stage in our real-world study

could be *missing not at random* (MNAR), requiring alternative approaches to multiple imputation. Since the MNAR assumption is never testable with observed data, the standard approach is to perform sensitivity analyses to different missingness mechanisms.^{27,2}

SIA and SMCFCFS imputation approaches give consistent estimation for logistic regression models with an interaction term when data are MAR, and either can be used in most analyses. SMCFCFS performed better than SIA when the fully observed variable in the interaction has an underlying continuous form. Passive imputation performed better than JAV imputation but these approaches should only be used (instead of SMCFCFS or SIA) when the coefficient of interest is the fully observed variable that has a low prevalence.

Author contributions

MJS, ENN and MQ developed the concept. MJS and MQ designed the first draft of the article and the computing code. All authors interpreted and reviewed the code and the data, and drafted and revised the article. All authors read and approved the final version of the article. MJS is the guarantor of the article.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Cancer Research UK (Reference C7923/A18525). Funders had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

Ethical statement

Ethical approval

We obtained the statutory approvals required for this research from the Confidentiality Advisory Group (CAG) of the Health Research Authority (HRA): PIAG 1–05(c) 2007. Ethical approval was obtained from the Research Ethics Committee (REC) of the Health Research Authority (HRA): 07/MRE01/52.

Informed consent

Informed consent from participants was waived by the ethics committee. This work uses the data provided by patients and collected by the National Health Service as part of their care and support. We used anonymized National Cancer Registry and Hospital Episode Statistics data. No consent to participate was sought from patients. All methods were carried out in accordance with relevant guidelines and regulations.

ORCID iD

Matthew J Smith  <https://orcid.org/0000-0002-8502-0056>

References

1. Rubin DB. *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley, 1987.
2. Carpenter JR and Kenward MG. *Multiple imputation and its application*. 1st edition. Hoboken, NJ: John Wiley & Sons, Ltd, 2013.
3. Von Hippel PT. How to impute interactions, squares, and other transformed variables. *Socio Methodol* 2009; 39(1): 265–291. DOI: [10.1111/J.1467-9531.2009.01215.X](https://doi.org/10.1111/J.1467-9531.2009.01215.X)
4. Seaman SR, Bartlett JW and White IR. Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods. *BMC Med Res Methodol* 2012; 12(1): 1–13. DOI: [10.1186/1471-2288-12-46/TABLES/5](https://doi.org/10.1186/1471-2288-12-46/TABLES/5)
5. Bartlett JW, Seaman SR, White IR, et al. Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. *Stat Methods Med Res* 2014; 24(4): 462–487. DOI: [10.1177/0962280214521348](https://doi.org/10.1177/0962280214521348)
6. Smith MJ, Njagi EN, Belot A, et al. Association between multimorbidity and socioeconomic deprivation on short-term mortality among patients with diffuse large B-cell or follicular lymphoma in England: a nationwide cohort study. *BMJ Open* 2021; 11(11): e049087. DOI: [10.1136/BMJOPEN-2021-049087](https://doi.org/10.1136/BMJOPEN-2021-049087)
7. Smith MJ, Belot A, Quartagno M, et al. Excess mortality by multimorbidity, socioeconomic, and healthcare factors, amongst patients diagnosed with diffuse large B-cell or follicular lymphoma in England. *Cancers* 2021 13(22): 5805. DOI: [10.3390/CANCERS13225805](https://doi.org/10.3390/CANCERS13225805)
8. Smith MJ, Fernandez MAL, Belot A, et al. Investigating the inequalities in route to diagnosis amongst patients with diffuse large B-cell or follicular lymphoma in England. *Br J Cancer* 2021 125: 1299. DOI: [10.1038/s41416-021-01523-6](https://doi.org/10.1038/s41416-021-01523-6)
9. Tang L, Song J, Belin TR, et al. A comparison of imputation methods in a longitudinal randomized clinical trial. *Stat Med* 2005; 24(14): 2111–2128. DOI: [10.1002/SIM.2099](https://doi.org/10.1002/SIM.2099)
10. Little RJ and Rubin DB. *Statistical analysis with missing data*. New York, NY: Wiley, 1987. DOI: [10.2307/1165119](https://doi.org/10.2307/1165119)
11. Buuren vS and Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. *J Stat Software* 2011; 45(3): 1–67. DOI: [10.18637/jss.v045.i03](https://doi.org/10.18637/jss.v045.i03)
12. Burton A, Altman DG, Royston P, et al. The design of simulation studies in medical statistics. *Stat Med* 2006; 25(24): 4279–4292. DOI: [10.1002/SIM.2673](https://doi.org/10.1002/SIM.2673)
13. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chron Dis* 1987; 40(5): 373–383. DOI: [10.1016/0021-9681\(87\)90171-8](https://doi.org/10.1016/0021-9681(87)90171-8)

14. Porta M. *A dictionary of epidemiology*. Oxford, UK: Oxford University Press, 2014. DOI: [10.1093/acref/9780195314496.001.0001](https://doi.org/10.1093/acref/9780195314496.001.0001)
15. Feinstein AR. The pre-therapeutic classification of comorbidity in chronic disease. *J Chron Dis* 1970; 23(7): 455–468. DOI: [10.1016/0021-9681\(70\)90054-8](https://doi.org/10.1016/0021-9681(70)90054-8)
16. NHS digital. Hospital episode statistics. 2015. DOI: <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics>
17. International Agency for Research on Cancer. *International classification of diseases for oncology, 3rd Edition (ICD-O-3)*. World Health Organisation, 2013. <https://www.who.int/standards/classifications/other-classifications/international-classification-of-diseases-for-oncology>
18. Maringe C, Fowler H, Rachet B, et al. Reproducibility, reliability and validity of population-based administrative health data for the assessment of cancer non-related comorbidities. *PLoS One* 2017; 12(3): e0172814. DOI: [10.1371/journal.pone.0172814](https://doi.org/10.1371/journal.pone.0172814)
19. Renzi C, Kaushal A, Emery J, et al. Comorbid chronic diseases and cancer diagnosis: disease-specific effects and underlying mechanisms. *Nat Rev Clin Oncol* 2019; 16(12): 746–761. DOI: [10.1038/s41571-019-0249-6](https://doi.org/10.1038/s41571-019-0249-6)
20. Smith MJ, Rachet B and Luque-Fernandez MA. Mediating effects of diagnostic route on the comorbidity gap in survival of patients with diffuse large B-cell or follicular lymphoma in England. *Cancers* 2022 14(20): 5082. DOI: [10.3390/CANCERS14205082](https://doi.org/10.3390/CANCERS14205082)
21. Seaman SR and White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res* 2013; 22(3): 278–295. DOI: [10.1177/0962280210395740/ASSET/IMAGES/LARGE/10.1177_0962280210395740-FIG1](https://doi.org/10.1177/0962280210395740/ASSET/IMAGES/LARGE/10.1177_0962280210395740-FIG1)
22. Li L, Shen C, Li X, et al. On weighting approaches for missing data. *Stat Methods Med Res* 2013; 22(1): 14–30. DOI: [10.1177/0962280211403597/ASSET/IMAGES/LARGE/10.1177_0962280211403597-FIG1](https://doi.org/10.1177/0962280211403597/ASSET/IMAGES/LARGE/10.1177_0962280211403597-FIG1)
23. White IR, Daniel R and Royston P. Avoiding bias due to perfect prediction in multiple imputation of incomplete categorical variables. *Comput Stat Data Anal* 2010; 54(10): 2267–2275. DOI: [10.1016/J.CSDA.2010.04.005](https://doi.org/10.1016/J.CSDA.2010.04.005)
24. Firth D. Bias reduction of maximum likelihood estimates. *Biometrika* 1993; 80(1): 27–38. DOI: [10.1093/BIOMET/80.1.27](https://doi.org/10.1093/BIOMET/80.1.27)
25. Kosmidis I and Firth D. Jeffreys-prior penalty, finiteness and shrinkage in binomial-response generalized linear models. *Biometrika* 2021; 108(1): 71–82. DOI: [10.1093/BIOMET/ASAA052](https://doi.org/10.1093/BIOMET/ASAA052)
26. Project TINHLPF. A predictive model for Aggressive non-hodgkin's lymphoma. *N Engl J Med* 1993; 329(14): 987–994. DOI: [10.1056/NEJM199309303291402](https://doi.org/10.1056/NEJM199309303291402)
27. Gachau S, Quartagno M, Njagi EN, et al. Handling missing data in modelling quality of clinician-prescribed routine care: sensitivity analysis of departure from Missing at Random (MAR) assumption. *Stat Methods Med Res* 2020; 29(10): 3076. DOI: [10.1177/0962280220918279](https://doi.org/10.1177/0962280220918279)

Appendix

A Tables

Table A1. Performance measures (relative bias, coverage, and relative error) of each imputation method in estimating the coefficients (fully observed, partially observed, and interaction), by data-generating mechanism.

| | Fully observed (Z) | | | | | | Partially observed (X) | | | | | | Interaction (XZ) | | | | | | | | | | | |
|----------------|--------------------|-------|------|-------|---------|-------|------------------------|-------|------|-------|------|-------|------------------|------|-------|-------|--------|-------|----------|---|------|---|---|--|
| | R. Bias (%) | | MCSE | | Cov | | R. Error | | MCSE | | Cov | | R. Bias (%) | | MCSE | | Cov | | R. Error | | MCSE | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | |
| No interaction | | | | | | | | | | | | | | | | | | | | | | | | |
| Passive | 0.0 | 0.003 | 93.4 | 0.008 | 2.149 | 0.004 | 0.0 | 0.004 | 94.7 | 0.007 | 1.7 | 2.294 | - | - | - | - | - | - | - | - | - | - | - | |
| JAV | 0.0 | 0.003 | 93.4 | 0.008 | 2.149 | 0.004 | 0.0 | 0.004 | 94.7 | 0.007 | 1.7 | 2.294 | - | - | - | - | - | - | - | - | - | - | - | |
| SIA | 0.0 | 0.003 | 93.7 | 0.008 | 2.149 | 0.004 | 0.0 | 0.004 | 94.9 | 0.007 | 1.5 | 2.289 | - | - | - | - | - | - | - | - | - | - | - | |
| SMCFCS | 0.0 | 0.003 | 93.6 | 0.008 | 2.149 | 0.004 | 0.0 | 0.004 | 95.1 | 0.007 | 0.9 | 2.276 | - | - | - | - | - | - | - | - | - | - | - | |
| DGMI | | | | | | | | | | | | | | | | | | | | | | | | |
| Passive | -50.7 | 0.004 | 95.6 | 0.007 | 2.857 | 0.004 | 4.5 | 0.004 | 96.0 | 0.006 | 6.1 | 2.440 | -44.3 | 0.00 | 95.9 | 0.006 | 61.1 | 2.993 | | | | | | |
| JAV | -162.4 | 0.004 | 75.4 | 0.014 | 2.960 | 0.004 | -3.7 | 0.004 | 95.5 | 0.007 | 3.7 | 2.421 | -119.5 | 0.01 | 81.8 | 0.012 | 52.5 | 4.277 | | | | | | |
| SIA | 8.1 | 0.005 | 95.2 | 0.007 | 2.324 | 0.004 | -0.5 | 0.004 | 94.9 | 0.007 | -0.3 | 2.297 | -1.1 | 0.01 | 94.1 | 0.008 | 0.7 | 1.713 | | | | | | |
| SMCFCS | 6.0 | 0.005 | 95.1 | 0.007 | 2.338 | 0.004 | -0.9 | 0.004 | 94.6 | 0.007 | 0.4 | 2.310 | -2.5 | 0.01 | 94.0 | 0.008 | 2.1 | 1.724 | | | | | | |
| DGM2 | | | | | | | | | | | | | | | | | | | | | | | | |
| Passive | -13.4 | 0.004 | 98.2 | 0.004 | 2.837 | 0.004 | 0.9 | 0.004 | 96.6 | 0.006 | 6.0 | 2.440 | -118.3 | 0.00 | 99.4 | 0.002 | 57.5 | 3.413 | | | | | | |
| JAV | -109.8 | 0.004 | 82.8 | 0.012 | 2.985 | 0.004 | -3.6 | 0.004 | 96.3 | 0.006 | 3.6 | 2.426 | -478.4 | 0.01 | 86.4 | 0.011 | 53.7 | 5.181 | | | | | | |
| SIA | 8.5 | 0.005 | 95.8 | 0.006 | 2.319 | 0.004 | -0.7 | 0.004 | 94.6 | 0.007 | -1.2 | 2.274 | 454.2 | 0.01 | 94.9 | 0.007 | -0.4 | 2.072 | | | | | | |
| SMCFCS | 7.2 | 0.005 | 95.1 | 0.007 | 2.330 | 0.004 | -0.6 | 0.004 | 94.6 | 0.007 | -0.4 | 2.292 | 100.6 | 0.01 | 94.7 | 0.007 | 0.8 | 2.074 | | | | | | |
| DGM3 | | | | | | | | | | | | | | | | | | | | | | | | |
| Passive | -118.0 | 0.004 | 83.7 | 0.012 | 2.892 | 0.004 | 11.6 | 0.004 | 93.6 | 0.008 | 4.1 | 2.395 | -103.3 | 0.00 | 83.4 | 0.006 | 56.4 | 2.582 | | | | | | |
| JAV | -249.3 | 0.004 | 56.2 | 0.016 | 3.062 | 0.004 | -4.0 | 0.004 | 95.8 | 0.006 | 2.4 | 2.427 | -168.0 | 0.01 | 67.7 | 0.008 | 46.8 | 3.729 | | | | | | |
| SIA | 6.2 | 0.005 | 95.3 | 0.007 | 2.323 | 0.004 | -0.9 | 0.004 | 94.5 | 0.007 | -2.0 | 2.257 | -0.5 | 0.01 | 95.5 | 0.006 | 1.3 | 1.313 | | | | | | |
| SMCFCS | 6.6 | 0.005 | 96.0 | 0.006 | 2.316 | 0.004 | -1.5 | 0.004 | 94.6 | 0.007 | -2.1 | 2.253 | -64.0 | 0.01 | 94.9 | 0.007 | 0.3 | 1.339 | | | | | | |
| DGM4 | | | | | | | | | | | | | | | | | | | | | | | | |
| Passive | -23.1 | 0.004 | 95.1 | 0.007 | 2.519 | 0.003 | 4.2 | 0.003 | 96.6 | 0.006 | 5.8 | 2.426 | -22.5 | 0.01 | 96.8 | 0.006 | 21.6 | 2.193 | | | | | | |
| JAV | -57.6 | 0.004 | 87.9 | 0.010 | 2.354 | 0.004 | -1.3 | 0.004 | 96.2 | 0.007 | 1.0 | 2.317 | -24.8 | 0.01 | 92.6 | 0.008 | 8.2 | 2.010 | | | | | | |
| SIA | 2.5 | 0.004 | 94.8 | 0.007 | 2.294 | 0.004 | -0.6 | 0.004 | 95.2 | 0.007 | 1.0 | 2.317 | -2.7 | 0.01 | 95.8 | 0.006 | 0.5 | 1.699 | | | | | | |
| SMCFCS | 2.8 | 0.004 | 95.1 | 0.007 | 2.290 | 0.004 | -0.6 | 0.004 | 95.3 | 0.007 | -0.5 | 2.282 | -2.4 | 0.01 | 95.3 | 0.007 | 0.3 | 1.702 | | | | | | |
| DGM5 | | | | | | | | | | | | | | | | | | | | | | | | |
| Passive | -46.8 | 0.003 | 95.6 | 0.007 | 2.856 | 0.003 | -6.7 | 0.003 | 94.5 | 0.007 | 1.7 | 2.348 | -38.3 | 0.00 | 95.9 | 0.006 | 46.6 | 2.785 | | | | | | |
| JAV | -22.6 | 0.004 | 94.7 | 0.007 | 2.394 | 0.003 | -3.7 | 0.003 | 94.3 | 0.007 | 1.5 | 2.342 | -13.8 | 0.01 | 93.7 | 0.008 | -1.8 | 1.752 | | | | | | |
| SIA | -68.3 | 0.001 | 62.3 | 0.015 | 2.437 | 0.008 | -77.1 | 0.008 | 75.3 | 0.014 | 1.2 | 2.329 | -68.0 | 0.00 | 35.6 | 0.015 | 1.9 | 2.021 | | | | | | |
| SMCFCS | -1.8 | 0.004 | 95.1 | 0.007 | 2.398 | 0.003 | -1.6 | 0.003 | 94.8 | 0.007 | 1.9 | 2.351 | -1.2 | 0.01 | 94.6 | 0.007 | -0.2 | 1.691 | | | | | | |
| DGM6 | | | | | | | | | | | | | | | | | | | | | | | | |
| Passive | 662.5 | 0.067 | 98.1 | 0.004 | 140.363 | 0.003 | -0.3 | 0.003 | 94.4 | 0.007 | 0.6 | 2.320 | -106.6 | 0.09 | 100.0 | 0.000 | 2757.1 | 3.477 | | | | | | |
| JAV | 638.7 | 0.052 | 99.4 | 0.002 | 173.719 | 0.003 | -1.5 | 0.003 | 95.0 | 0.007 | 1.0 | 2.333 | -237.5 | 0.09 | 100.0 | 0.000 | 2949.0 | 4.877 | | | | | | |
| SIA | 1925.6 | 0.123 | 97.1 | 0.005 | 85.316 | 0.003 | -0.8 | 0.003 | 95.0 | 0.007 | 1.2 | 2.334 | 101.3 | 0.18 | 99.5 | 0.002 | 1995.8 | 1.496 | | | | | | |
| SMCFCS | 1298.1 | 0.085 | 97.8 | 0.005 | 110.709 | 0.003 | -0.6 | 0.003 | 95.1 | 0.007 | -0.1 | 2.305 | -27.1 | 0.12 | 99.5 | 0.002 | 2774.8 | 1.510 | | | | | | |

R. Bias: relative bias, Cov.: coverage of 95% confidence intervals, R. Error: relative error of model-based standard error and empirical error, MCSE: Monte-Carlo Standard Error, JAV: Just Another Variable, SIA: Stratify-Impute-Append, SMCFCFS: substantive model compatible fully conditional specification, DGM: data-generating mechanism.