

RESEARCH

Open Access



# Comparison of different approaches in handling missing data in longitudinal multiple-item patient-reported outcomes: a simulation study

Minqian Yan<sup>1</sup>, Lizhi Zhou<sup>1</sup>, Chongye Zhao<sup>1</sup>, Chen Shi<sup>1</sup> and Chunquan Ou<sup>1,2\*</sup>

## Abstract

**Background** Patient-reported outcomes (PROs) are important clinical outcomes widely used as primary and secondary endpoints in clinical studies. However, PRO data often suffers from missing values for various reasons, which pose challenges to data analysis. This simulation study aimed to compare the performance of existing state-of-the-art approaches in handling missing PRO data.

**Methods** Using a real and complete multiple-item PRO dataset, we generated various missing scenarios with different missing rates, mechanisms, and patterns. The performances of eight methods were compared, including a mixed model for repeated measures (MMRM) with and without imputation at the item level, multiple imputation by chained equations (MICE) at the composite score and item levels, and three control-based pattern mixture models (PPMs) and the last observation carried forward (LOCF) imputation at the item level.

**Results** We found that the bias (i.e., deviation of the estimated from the true value) in the treatment effect estimates increased, and the statistical power diminished as the missing rate increased, especially for monotonic missing data. Item-level imputation led to a smaller bias and less reduction in power than composite score-level imputation. Except for cases under missing-not-at-random mechanisms (MNAR) and with a high proportion of patients' entire questionnaire missing, MMRM imputation at the item level demonstrated the lowest bias and highest power, followed by MICE imputation at the item level. The PPM methods were superior to the other methods under MNAR mechanisms.

**Conclusions** PPMs imputation at the item level was preferable for MNAR, whereas MMRM and MICE imputation at the item level were better for other scenarios. These findings provide valuable insight for selecting appropriate methods for handling missing PRO data.

**Keywords** Patient-reported outcomes, Mixed model for repeated measures, Multiple imputations, Pattern mixture models, Randomised controlled trials, Monte carlo simulation

\*Correspondence:  
Chunquan Ou  
ouchunquan@hotmail.com

<sup>1</sup>State Key Laboratory of Organ Failure Research, Department of  
Biostatistics, Guangdong Provincial Key Laboratory of Tropical Disease

Research, School of Public Health, Southern Medical University,  
Guangzhou 510515, Guangdong, China

<sup>2</sup>Department of Biostatistics, School of Public Health, Southern Medical  
University, Guangzhou 510515, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Background

Patient-reported outcomes (PROs), an important form of clinical outcomes, are directly reported by patients themselves and cannot be modified or interpreted by others [1, 2]. As the concept and practice of patient-focused drug development (PFDD) continue to evolve, PROs have garnered increasing attention and are frequently used as the primary or secondary endpoint in clinical trials [3, 4]. The most common PRO measures are scales, particularly health-related quality-of-life (HRQL) instruments.

However, missing PRO data is a prevalent issue in clinical trials [5, 6]. Patients may fail to report all PRO measures or omit specific items in an instrument. Missing data poses significant challenges to data analysis, potentially increasing the standard error, reducing statistical power [3–10], introducing bias in the estimates of treatment effects [11], and ultimately affecting the scientific integrity and value of conclusions. Despite the critical importance of handling missing data, many studies have overlooked this issue in PROs. A recent review showed that 18% of the trials using PROs as the primary endpoints did not report the missing data rate, and only 7% described the statistical methods for handling missing data, with 75% relying on single imputation methods. Less than 10% of the trials conducted sensitivity analyses to justify their approaches to handling missing data [12].

Selecting appropriate approaches for handling missing data is a complicated and vital task that can significantly influence the validity of results. Various statistical methods exist, each with specific assumptions and limitations [13]. The main methods include single imputation methods (e.g., last observation carried forward, LOCF), model-based multiple imputation methods [14, 15], maximum likelihood modeling approaches, and various sensitivity analyses. The improper selection of these methods can lead to biased estimates and misleading conclusions. Therefore, it is essential to summarize the applicability and particularities of different imputation methods and discuss their robustness.

In many clinical studies on non-PRO outcomes, single imputation methods, especially LOCF, have been widely used across various therapeutic areas [16]. However, the problems associated with single imputation are well documented [17–19], since they underestimate the uncertainties associated with the missing data and are more likely to bias the estimation of treatment effects compared to multiple imputation methods. Some simulation studies on non-PRO outcomes showed that LOCF can increase Type I error rates [20–22]. Nevertheless, we need to evaluate the performance of LOCF as a simple, straightforward, and common method for handling complicated PRO missing data in comparison to other multiple imputation methods. For general longitudinal data, model-based multiple imputation (MI) is recommended,

particularly for non-monotonic missing data [20, 23, 24]. One of the most commonly used MI techniques is multiple imputation by chained equations (MICE) [25–27]. Other methods, such as the mixed model for repeated measures (MRMM) [28–30] and pattern mixture models (PPMs) [31], have also been widely used in different studies. PPMs are frequently utilized for handling MNAR missing data in longitudinal clinical trials with continuous and binary outcomes [32–36], including Jump to Reference (J2R), Copy Reference (CR), and Copy Increments in Reference (CIR) [37], which are particularly used in trials for new drug development as stated in regulatory guidelines by FDA [38] and EMA [39]. These approaches, which impute missing values in the treatment group using models from the control group, may offer a conservative estimate of treatment effects [36]. In non-PRO outcome studies, J2R is the most conservative among all PPM variants [36]. The CR method is also conservative but less so than J2R, as it incorporates carry-over treatment effects by using prior observed values in the active treatment group as predictors [37]. PPMs do not inflate Type-I errors under MAR. Simulations indicate that Type-I error inflation for PPMs is comparable to that for MMRM when data deviates from MAR [36].

Longitudinal PRO data present unique challenges and opportunities for handling missing data due to the correlation between repeated measures over time and between intra-instrument items at the same time point. Additionally, missing data sometimes occurs for specific items (i.e., item non-response) rather than for all PRO measures (i.e., unit non-response) [40, 41]. However, studies comparing the performance of different imputation methods in longitudinal clinical trials with a multiple-item PRO are lacking. Research addressing these issues can provide insight into the handling of missing data on PRO measures in longitudinal studies. Some studies have suggested that MI at the item level is advantageous for statistical accuracy, especially when the sample size is less than 500 and the missing data rate is over 10% [42, 43]. However, these studies simulated data from a single visit by using a fixed proportion of unit non-response. It is still unclear whether item-level, dimension-level, or composite score-level imputation should be selected for multiple-visit PROs and which method is superior under monotonic or non-monotonic missing data with different proportions of unit non-response. Missing pattern, including monotonic and non-monotonic missing, is another issue that needs to be considered when comparing different imputation methods. Monotone missing refers to the fact that once a participant has data missing at a time point, all subsequent measurements of that participant are also missing. This is a common phenomenon caused by participant drop-outs in longitudinal clinical trials [44]. Non-monotonic missing means that a

participant has data missing at a time point but present at subsequent time points [44]. Prior research on non-PRO outcomes suggests that MI is more effective in handling non-monotonic missing data compared to monotonic cases [44, 45].

This study simulated various missing scenarios with different patterns and rates of missing data based on a real and complete PRO dataset. Eight common methods, including a direct mixed model for repeated measures (MMRM), single imputation, and multiple imputation at the composite score and item levels, were used to handle the missing data, and their statistical performances were compared. This study aimed to provide a simulation-based evaluation of the advantages and disadvantages of different methods under various scenarios. The results would offer valuable suggestions for choosing appropriate methods for handling missing PRO data, thereby enhancing the integrity and reliability of the analytical results.

## Methods

### Clinical trial data

This study utilized real data from a randomized placebo-controlled double-blind trial on depression that included 180 Chinese patients primarily with depression, anxiety, and insomnia. In this trial, patients were randomized into a control group receiving a placebo and a treatment group receiving traditional Chinese medicine. The primary outcome was measured using the 17-item Hamilton Depression Scale (HAMD-17). The item and scale scores were repeatedly measured at baseline, 2, 4, and 6 weeks of follow-up. The primary outcome was the change from baseline at week 6. Baseline covariates, such as age and sex, were also assessed. The original complete dataset was analysed using an MMRM, and the estimate of the treatment effect was regarded as the “true” effect. All of the missing scenarios were simulated based on the original complete dataset.

### Methods to be compared

To focus on methods that are commonly used, practical, and widely accepted by regulatory authorities in the context of confirmatory RCTs, we primarily compared four types of methods, which were described as follows, with more detailed information provided in the Supplement.

#### *Mixed model for repeated measures (MMRM)*

An MMRM can analyse all cases, including missing values, without the need for imputation. The model parameters of population distribution were estimated using the maximum likelihood method. MMRM utilizes all information from all repeated measures, including the subjects with missing data, without the need for missing imputation [46, 47]. In this study, MMRM was used both

for direct analysis without imputation and predictive imputation. For the direct analysis, MMRM included the treatment group, visit time, and their interaction as fixed effects, and within-subject variability as random effects, after adjusting for the fixed covariate factors (e.g., age and gender). For the predictive imputation, we conducted the MMRM direct analysis and used this model to impute the missing data. After performing missing data imputation using all the methods in this study, we also applied the MMRM to estimate the treatment effects. The results were then compared to the true value derived from the MMRM direct analysis based on the real complete data, and the true effect was defined as the coefficient of the treatment group variable in the MMRM direct analysis.

#### *Pattern mixture models (PPMs)*

The idea behind the PPMs is to construct a joint probability distribution of the observed and missing data, which is related to the missing probability distribution under the condition of observed covariates [32, 35, 48, 49]. The control-based PPMs in this study include jump-to-reference (J2R), copy reference (CR), and copy increment from reference (CIR) methods that impute missing data at the item level. PPMs are implemented based on multiple imputations.

#### *Multiple imputation by chained equations (MICE)*

MICE imputes the missing data in a dataset through a series of iterations of the prediction model [50]. In each iteration, the other variables in the dataset are used to estimate each specified variable in that dataset. These iterations are performed continuously until convergence is achieved. The imputation of MICE is highly flexible and can account for statistical uncertainty. In this study, an incomplete data set was imputed five times, and then each of the five imputed datasets was analysed using MMRM separately, and the analysis results were combined according to Rubin's rule [51].

#### *Last observation carried forward (LOCF)*

LOCF is a frequently used single imputation method that imputes missing values using the last observation at the item level [52]. In this study, LOCF, as a representative of single imputation methods, was compared with multiple imputation methods.

Imputation at the composite level for longitudinal data is equivalent to that for general longitudinal data. Since MICE using the predicted mean matching method is the preferred imputation approach for longitudinal data [20, 23]. Consequently, eight different methods were used to process the missing data and were compared: direct MMRM analysis (without imputation) and MICE using the predicted mean matching method imputation at the composite-score level and J2R, CR, CIR, MICE, MMRM,

and LOCF imputation at the item level. After imputation using other methods, MMRM was also used to estimate the effect.

### Missing mechanisms

There are three missing mechanisms [53]: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Data can be considered MCAR if the probability of missing data is independent of any observed or unobserved factors. Data will be considered MAR if the probability of missing may depend on observed covariates describing patients' characteristics (e.g., missing probability varied by sex) [54, 55]. For the MNAR mechanism, missingness depends on unobserved measurements [56]. In this study, missingness depends on the outcome, the patient's HAMD-17 score changes, which are unobservable once missing but they are all known in our true complete dataset.

### Simulation studies

In longitudinal studies, the composite score of multi-item scales as common PRO measures presents both complicated within-subject correlation over time and between-item correlation at the same time point, which are difficult to simulate. Therefore, as in other simulation studies on scales [5, 45], based on a real complete dataset, we simulated 90 different missing scenarios, namely 90 combinations of four parameters: three missing mechanisms, five missing rates, monotonic missing or non-monotonic missing, and three proportions of unit missing and item missing (Table 1). The specifications of various parameters in our simulation were based on common settings in real research [45] and statistical considerations [57–60].

Some data were sampled from the real dataset and assigned as missing. Three missing mechanisms, MCAR, missing at random (MAR) with more missing among females, and missing not at random (MNAR), were simulated as follows: MCAR was generated by sampling completely randomly; MAR was generated by randomly sampling at a male-to-female ratio of 1:3. Specifically,

for males and females, the probabilities of missing were set to  $0.25 \times$  individual missing rate (specified as 5%, 10%, 15%, 20%, or 30%) and  $0.75 \times$  individual missing rate, respectively; and for MNAR, we generated missing data of HAMD-17 score changes by randomly sampling at a 1:3 ratio for those below and above the mean score changes of all participants, implying that patients with higher value of the outcome (i.e. greater HAMD-17 score change from baseline) are more likely to have missing data. The three missing mechanisms were applied to the data at five different missing rates: 5%, 10%, 15%, 20%, and 30%. The specific missing rate randomly determined which individuals would have data omitted, and then these individuals were randomly divided into two categories: unit non-response and item non-response, which were simulated in three different ratios (i.e., 0.2:0.8, 0.5:0.5, 0.8:0.2). The missing probabilities of the second-, fourth-, and sixth-week visits of missing individuals were set at 30%, 30%, and 40%, respectively. If an individual was classified as unit non-response in a certain visit time point, all the items in this time point would be missing; on the contrary, if an individual was classified as item non-response, only a randomly selected item in this time point would be missing. In the case of monotonic missing data, once a subject had missing data at one visit, all subsequent visit points were also missing.

Each of the 90 scenarios was simulated 1,000 times. Type I errors were generated by generating a binomial random variable with a probability of 0.5, which was uncorrelated with HAMD-17 score changes (i.e., the outcome). Then, we tested the null hypothesis that the coefficient of this variable in the MMRM model is equal to 0. Among 1,000 simulations, the proportion of simulations with a p-value less than 0.05 (i.e., rejecting the null hypothesis) was the type I error [61]. Power was the proportion of simulations with the observed p-value of the treatment less than 0.05 under the alternative hypothesis.

The performances of different methods for handling missing data were assessed using the mean absolute error (MAE), mean relative error (MRE), root mean square error (RMSE), type I error, and statistical power. In simulation studies, MAE, MRE, and RMSE are common performance indicators used to evaluate the deviation of estimates from the true values [5, 45]. The equations for MAE, MRE, and RMSE are as follows.

$$MAE = \text{mean}(|\tau - T|) \quad (1)$$

$$MRE = \text{mean}(|\tau - T|/\tau) \quad (2)$$

$$RMSE = \sqrt{\text{mean}((\tau - T)^2)} \quad (3)$$

**Table 1** Summary of parameters in simulating missing data

Parameters	Details
Missing rate	5%, 10%, 15%, 20%, 30%, on behalf of the proportion of individuals with missing items
The proportion of unit non-response and fixed item non-response	0.2:0.8, 0.5:0.5, 0.8:0.2, on behalf of the probability of two different missing patterns
Missing probabilities of the second-, fourth-, and sixth-week visits of the missing individuals	30%:30%:40%, baseline was not set for missing
Missing mechanisms	MCAR, MAR, MNAR
Monotonicity	Monotonic and non-monotonic

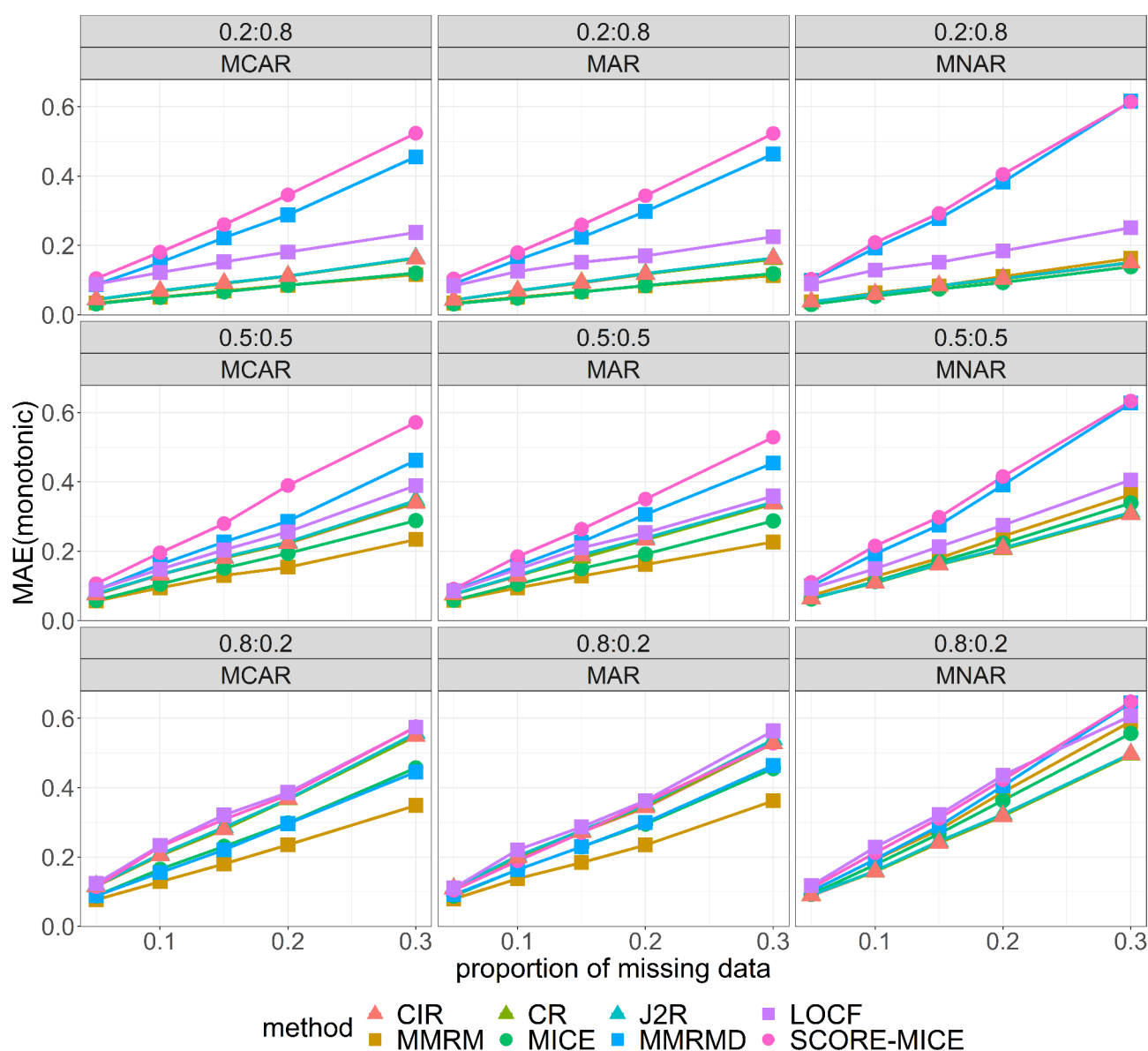
where  $T$  is the true effect of the treatment and  $\tau$  is its estimate. All simulations and analyses were performed using R 4.2.2 [62]. The R packages used for simulation and imputation in the study included “mmrm” and “mice”.

## Results

The original dataset included 180 patients (males: 90, females: 90). Their ages ranged from 24 to 68 years, with an average of 54.7 years (Standard deviation = 10.0). Descriptive statistics for the baseline covariates are

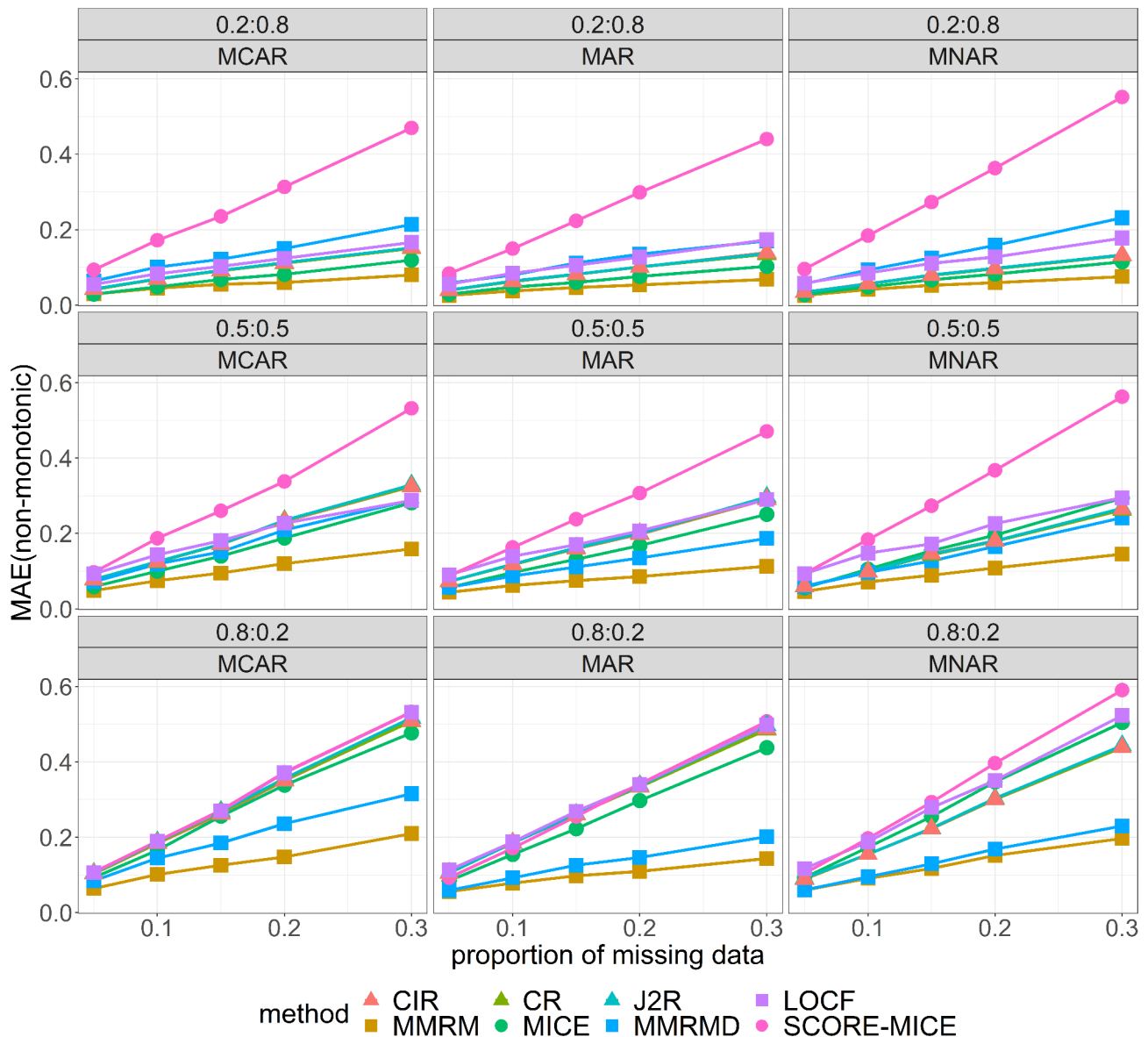
presented in Supplemental Table S1. The “true” effect in this study was  $-2.52$  ( $p < 0.001$ ).

Figures 1, 2 present the MAE for different imputation methods in the cases of monotonic and non-monotonic missing data, respectively. Unsurprisingly, the MAE for all methods increased with the missing rate. The performance of most methods was consistently worse under the MNAR mechanism compared to MCAR and MAR, with the difference increasing with the missing rate. Notably, under the three missing mechanisms, the MAE of each method in non-monotonic missing cases was



**Fig. 1** MAE in the monotonic missing cases (the proportion of the unit non-response and the item non-response includes 0.2:0.8, 0.5:0.5, 0.8:0.2). Abbreviations: MAE mean absolute error, MCAR missing completely at random, MAR missing at random, MNAR missing not at random, MMRM mixed model for repeated measures predictive imputation at the item level, MICE multiple imputation by chained equations at the item level, MMRMD mixed model for repeated measures directly at the composite score level, SCORE-MICE multiple imputation by chained equations at the composite score level, J2R jump to reference at the item level, CR copy reference at the item level, CIR copy increments in reference at the item level, LOCF last observation carried forward





**Fig. 2** MAE in the non-monotonic case (the proportion of the unit non-response and the item non-response includes 0.2:0.8, 0.5:0.5, 0.8:0.2). Abbreviations: MAE mean absolute error, MCAR missing completely at random, MAR missing at random, MNAR missing not at random, MMRM mixed model for repeated measures predictive imputation at the item level, MICE multiple imputation by chained equations at the item level, MMRMD mixed model for repeated measures directly at the composite score level, SCORE-MICE multiple imputation by chained equations at the composite score level, J2R jump to reference at the item level, CR copy reference at the item level, CIR copy increments in reference at the item level, LOCF last observation carried forward

smaller than that in monotonic cases. Similar results were observed for the MRE and RMSE (see Supplement for more details).

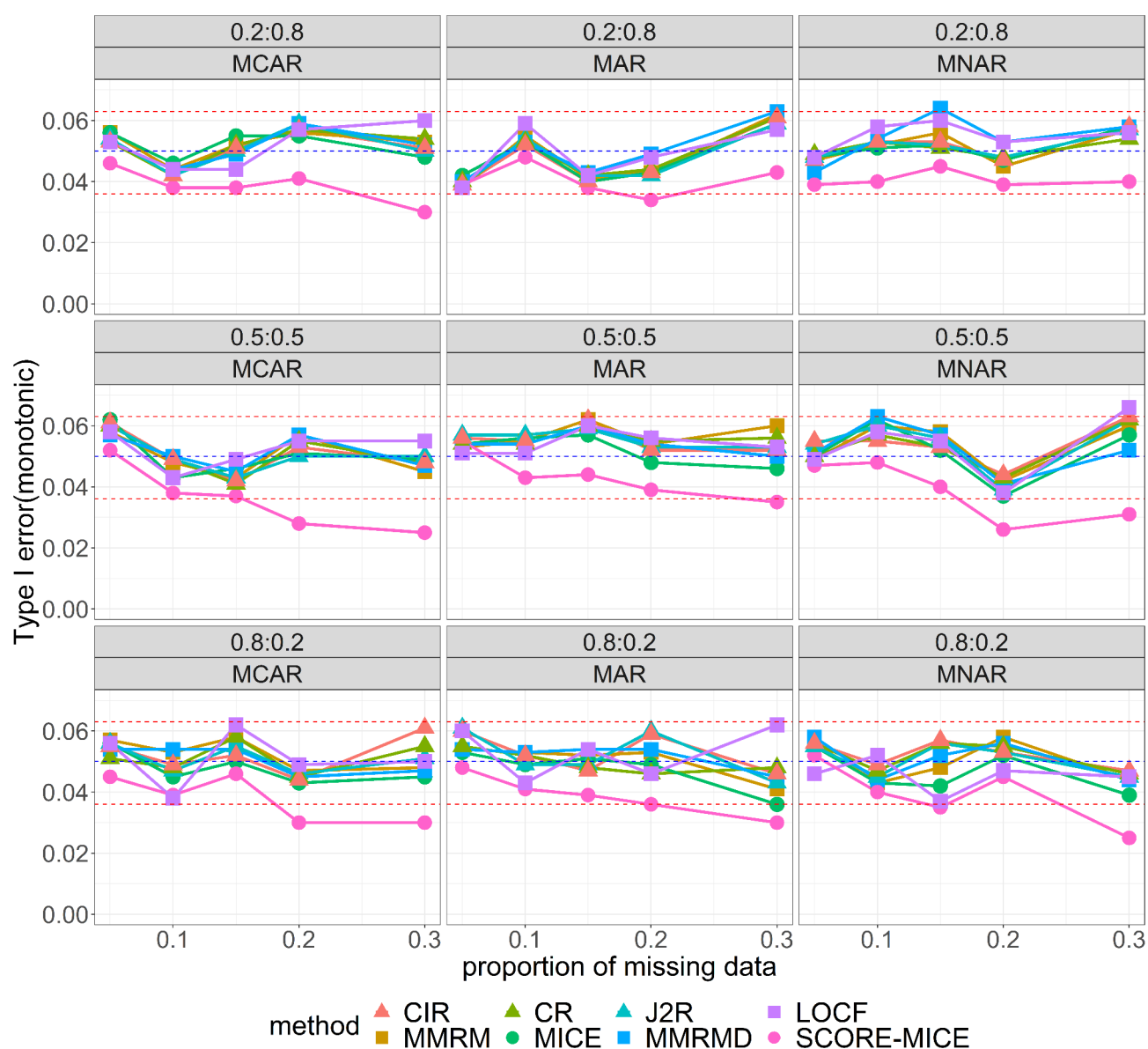
The performance of multiple imputations by MICE at the composite score level (SCORE-MICE) was inferior to other methods in both monotonic and non-monotonic missing cases, with the highest MAE, MRE, and RMSE under almost all scenarios. The performance difference between SCORE-MICE and other methods became smaller at a higher proportion of unit non-response. In monotonic missing cases, direct analysis without

imputation (MMRMD) also demonstrated a relatively high MAE, especially under the MNAR mechanism, and its MAE was higher than that of all other methods except for SCORE-MICE when the proportion of unit non-response was 0.2 and 0.5 in monotonic missing cases. Compared to other methods, MMRM with imputation at the item level performed the lowest MAE, except under the MNAR mechanism in monotonic missing cases. Similar results were observed in terms of the MRE and RMSE (see Supplement for more details). In monotonic missing cases, the MAE of MICE was lower than all other

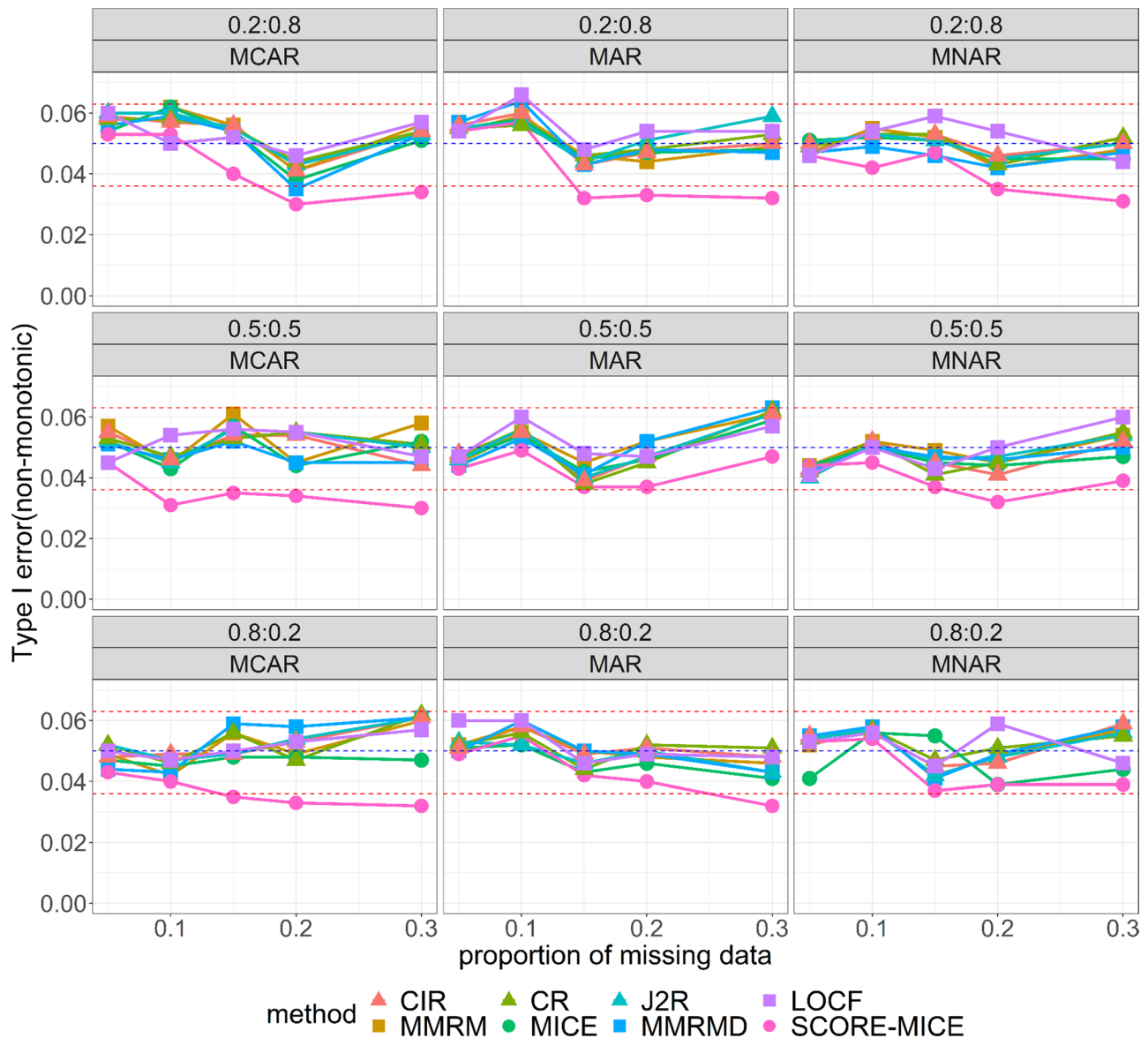
methods except for MMRM imputation at the item level under MCAR and MAR. PPM methods showed smaller MAE than other methods under MNAR. As the proportion of unit non-response increased (from 0.2 to 0.8), the difference among the various methods narrowed. The three PPM approaches (J2R, CR, and CIR) showed similar performances of MAE, MRE, and RMSE across scenarios and were superior to other methods under MNAR in monotonic cases, whereas the MMRM approach continued to perform well (i.e., lowest MAE, MRE, and RMSE) under MNAR in non-monotonic cases. The

LOCF approach was generally the worst (highest MAE, MRE, and RMSE) among the imputation methods at the item level (i.e., PPMs, MMRM, and LOCF) but better than SCORE-MICE and MMRMD when the proportion of unit non-response was 0.2 and 0.5.

Figures 3 and 4 show the simulation results for type I errors in the monotonic and non-monotonic cases, respectively. When the missing rate was greater than 20%, the SCORE-MICE tended to be conservative (type I error is less than the lower limit of the interval). Other



**Fig. 3** Type I errors in the monotonic case (the proportion of the unit non-response and the item non-response includes 0.2:0.8, 0.5:0.5, 0.8:0.2). Abbreviations: *MCAR* missing completely at random, *MAR* missing at random, *MNAR* missing not at random, *MMRM* mixed model for repeated measures predictive imputation at the item level, *MICE* multiple imputation by chained equations at the item level, *MMRMD* mixed model for repeated measures directly at the composite score level, *SCORE-MICE* multiple imputation by chained equations at the composite score level, *J2R* jump to reference at the item level, *CR* copy reference at the item level, *CIR* copy increments in reference at the item level, *LOCF* last observation carried forward



**Fig. 4** Type I error in the non-monotonic case (the proportion of the unit non-response and the item non-response includes 0.2:0.8, 0.5:0.5, 0.8:0.2). Abbreviations: *MCAR* missing completely at random, *MAR* missing at random, *MNAR* missing not at random, *MMRM* mixed model for repeated measures predictive imputation at the item level, *MICE* multiple imputation by chained equations at the item level, *MMRMD* mixed model for repeated measures directly at the composite score level, *SCORE-MICE* multiple imputation by chained equations at the composite score level, *J2R* jump to reference at the item level, *CR* copy reference at the item level, *CIR* copy increments in reference at the item level, *LOCF* last observation carried forward

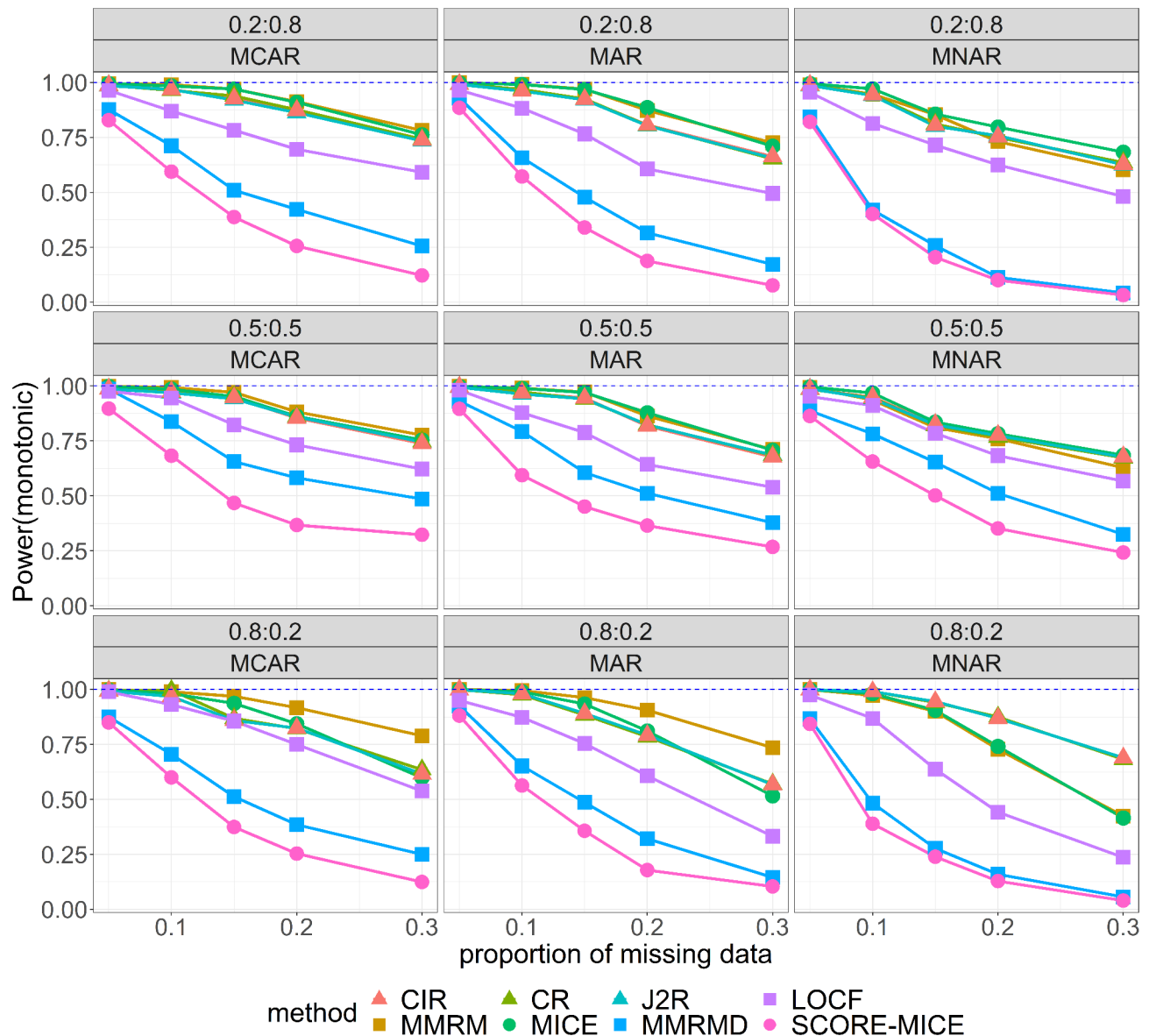
methods generally controlled for type-I errors in all scenarios

Figures 5 and 6 show the power results in the monotonic and non-monotonic missing cases, respectively. Undoubtedly, the power decreased as the missing rate increased. The MMRM, MICE, and PPMs methods exhibited remarkably higher power than the MMRMD and SCORE-MICE methods across various scenarios. Although the MAE, MRE, or RMSE performance of the MMRMD method may be good in non-monotonic terms, its power was relatively low both in the monotonic and non-monotonic missing cases. The power of

LOCF was intermediate across scenarios. The MICE, the PPMs, and the MMRM had higher power when the proportion of unit non-response was 0.2 or 0.5 both in the monotonic and non-monotonic missing cases. In addition, when the proportion of unit non-response was 0.5 or lower, the power of the MICE and PPMs approaches was higher than the other approaches. When the proportion of unit non-response was 0.8, the PPMs performed the best under MNAR (i.e., highest power).

The study integrated all simulation results and summarized the applicable scenarios for each method, which are shown in Table 2.





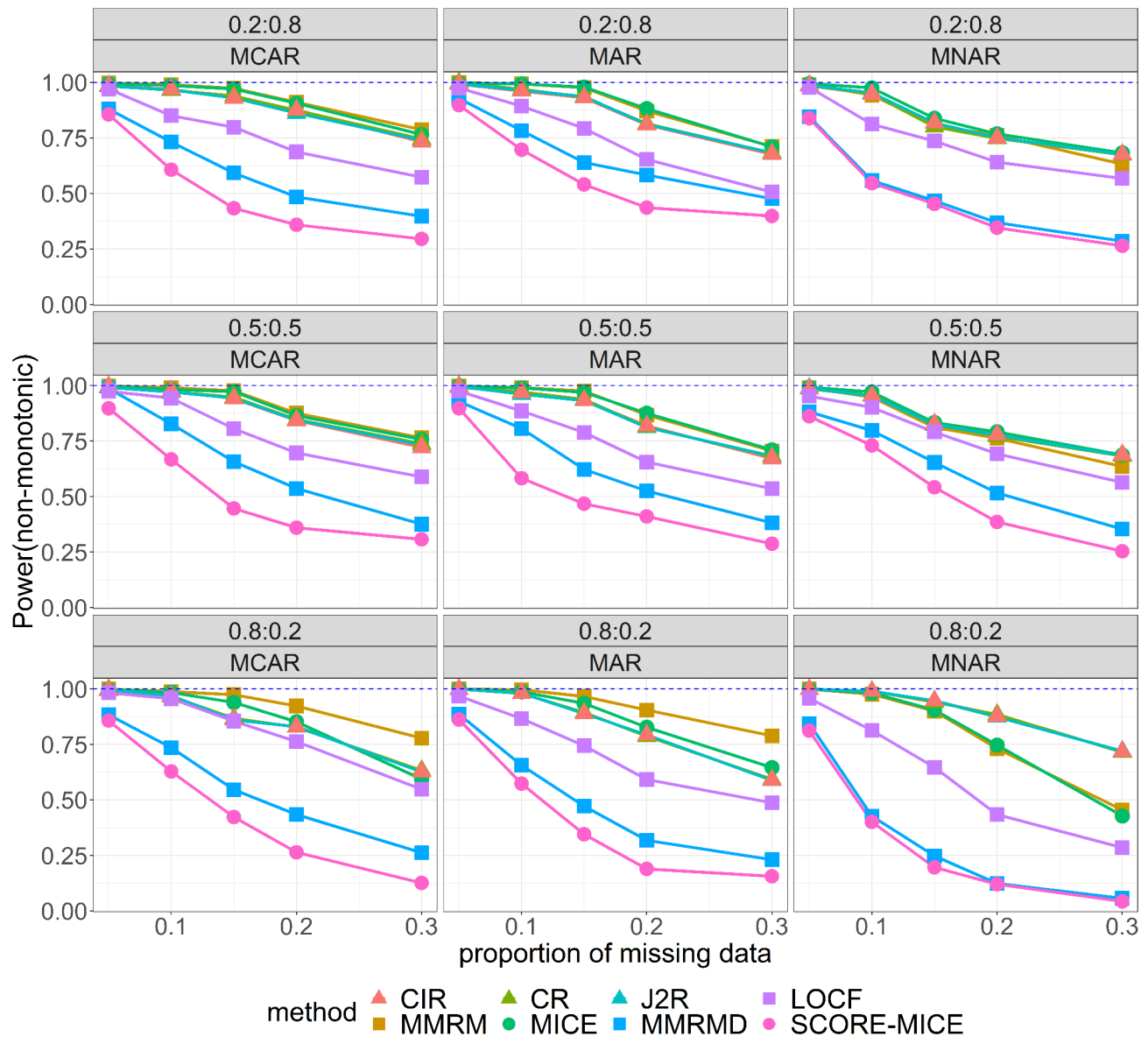
**Fig. 5** Power in the monotonic case (the proportion of the unit non-response and the item non-response includes 0.2:0.8, 0.5:0.5, 0.8:0.2). Abbreviations: *MCAR* missing completely at random, *MAR* missing at random, *MNAR* missing not at random, *MMRM* mixed model for repeated measures predictive imputation at the item level, *MICE* multiple imputation by chained equations at the item level, *MMRMD* mixed model for repeated measures directly at the composite score level, *SCORE-MICE* multiple imputation by chained equations at the composite score level, *J2R* jump to reference at the item level, *CR* copy reference at the item level, *CIR* copy increments in reference at the item level, *LOCF* last observation carried forward

## Discussion

Based on a real and complete RCT dataset with the HAM-D-17 as the primary outcome, this study simulated various missing scenarios with different missing patterns and rates. Our findings indicated that regardless of the type of missing mechanism, all methods are more accurate (lower MAE, MRE, and RMSE) in non-monotonic missing cases compared to monotonic cases. The statistical power significantly diminished as missing rates increased, especially under the MNAR mechanism. Item-level imputation approaches, with higher power

and lower bias, were superior to composite score-level imputation across scenarios, even with a high proportion of unit non-response. Specifically, for MNAR in the monotonic cases, PPMs (J2R, CR, and CIR) were preferable, whereas the MMRM and MICE were more suitable for MCAR and MAR in non-monotonic cases.

Several simulation studies have compared different approaches for handling missing data. However, they only considered a single visit. Moreover, PRO data tended to be incomplete in longitudinal follow-up studies [63, 64], in which repeated measurements at multiple follow-up



**Fig. 6** Power in the non-monotonic case (the proportion of the unit non-response and the item non-response includes 0.2:0.8, 0.5:0.5, 0.8:0.2). Abbreviations: *MCAR* missing completely at random, *MAR* missing at random, *MNAR* missing not at random, *MMRM* mixed model for repeated measures predictive imputation at the item level, *MICE* multiple imputation by chained equations at the item level, *MMRMD* mixed model for repeated measures directly at the composite score level, *SCORE-MICE* multiple imputation by chained equations at the composite score level, *J2R* jump to reference at the item level, *CR* copy reference at the item level, *CIR* copy increments in reference at the item level, *LOCF* last observation carried forward

**Table 2** Summary of the applicable scenarios for each method

method	monotonicity	missing mechanisms	the ratio of the non-response units and the non-response items
MMRM	monotonic	MCAR, MAR	Unrestricted
	non-monotonic	MCAR, MAR, MNAR	Unrestricted
MICE	monotonic	MCAR, MAR, MNAR	The proportion of unit non-response $\leq 0.5$
	non-monotonic	MCAR, MAR	The proportion of unit non-response $\leq 0.5$
PPMs (J2R, CR, CIR)	monotonic	MNAR	Unrestricted
	non-monotonic	-	-

SCORE-MICE, MMRMD, and LOCF methods are not recommended in any scenarios

visit points were common and presented a within-subject correlation [65]. This study addressed this gap by considering multiple visit points, monotonic and non-monotonic missing cases, and different proportions of unit and item non-response. The simulation was based on complete real longitudinal PRO data, ensuring that the data closely resembled real-world scenarios and maintained the relationship between items and visit points.

This study emphasizes the importance of item-level imputation in maintaining data integrity and analysis accuracy. Simons et al. [42] found that imputation at the item level of the EQ-5D-3 L outperformed the composite score-level imputation as the proportion of missing items increased under the MAR missing mechanism. Their study also showed that when the sample size was less than 500 and the missing rate was less than 10%, both imputation approaches yielded similar results for single-visit PRO data [42]. Similarly, Eekhout et al. reported better results with item-level imputation than with single imputation and MI on the pain coping inventory (PCI-active) in a single visit [43]. Our findings support the idea that item-level imputation generally performed better than composite score-level imputation for multiple-visit PRO data, although these two imputations were similar when the missing rate was low (<10%). The composite-level imputation is simple and commonly used, yet it is uncertain whether it outperforms item-level imputation. Our results demonstrated that even when a large proportion of items are missing, item-level imputation remains superior. Therefore, it is advisable to use as many individual items as possible, even if only a few are available.

Monotonic and non-monotonic missing may have different impacts on the performance of imputation methods. Rombach et al. indicated that MI performed better in the non-monotonic case than in the monotonic case for composite score-level imputation under MCAR and MAR missing mechanisms [45]. However, they did not consider item-level imputation, the MCAR and MNAR missing data mechanisms, nor did they take into account the PPMs, MMRM, and LOCF methods. Our study contributed significant evidence by demonstrating that most methods consistently achieved higher power and lower bias in non-monotonic cases than in monotonic cases under three different missing mechanisms for both item- and composite-level imputation. This difference is likely due to the higher amount of missing data in monotonic cases. In addition, we found that the three control-based PPM methods were relatively similar in terms of performance, probably because they shared many underlying principles. Under the MNAR mechanism, the PPMs generally perform well with the lowest MAE, MRE, and RMSE, particularly in monotonic cases. Its advantage under the MNAR missing mechanism in general longitudinal data has been proven previously [66, 67]. Under the

MAR mechanism, methods involving covariates, such as the MMRM, SCORE-MICE, and MICE, performed better with higher power and lower bias than under the MCAR mechanism. This improvement is likely because covariates were considered in the imputation process, and the MAR mechanism is related to covariates. Additionally, regarding the SCORE-MICE method, the RMSE results reported in this study are comparable to those reported by Rombach et al. (2018) [45]. As shown in the RMSE results from Rombach et al. (2018), all the RMSE results for 10%, 20%, and 30% missing rates in our study, with a sample size of 180, fall within the range of results obtained for sample sizes of 100 and 250.

In general, most approaches performed worse with lower power and higher bias when the proportion of unit non-responses was high because more items tended to be missing. Researchers should try to avoid missing PRO data during the trial design and implementation phases, such as training researchers, reviewing missing PRO scale data, and excluding potential subjects who are unlikely or unable to comply with follow-up in advance [67].

This study had several limitations. First, as in previous simulation studies on multi-item PROs, the simulation data were generated based on a real RCT dataset. Future studies using different PRO measures may validate our results. Second, the simulated scenarios were limited to specific missing data patterns and proportions of missing data. However, we believe that missing data rates between 5% and 30% are representative of the vast majority of RCTs. The missing patterns used were based on those observed in a common clinical trial. These patterns are realistic and representative. Although we simulated various missing scenarios, the real world is more complex and varied. Consequently, it remains uncertain whether the findings of this simulation study are fully applicable to all real-world settings. Our study prioritized methods that are most commonly used, practical, and widely accepted by regulatory authorities in the context of confirmatory RCTs. Future research could further explore and compare alternative methods, such as the expectation-maximization algorithm and the Bayesian method, to provide a more comprehensive understanding of handling PRO missing data in both clinical trials and observational studies.

## Conclusions

Our simulation study demonstrated that item-level imputation generally had higher power and accuracy than composite score-level imputation, even with a missing rate of up to 30%. Overall, the bias of all approaches increased with the proportion of unit non-response, and the power decreased as the missing rate increased. When selecting the imputation method, the PPM methods were more suitable for monotonic cases under the

MNAR mechanism since they demonstrated high power and small bias regardless of the proportion of unit non-response, whereas the MMRM and MICE were more suitable for MCAR and MAR in non-monotonic cases. These findings provide important insight for the selection of appropriate methods for handling missing PRO data.

#### Abbreviations

CIR	Copy increment from reference
CR	Copy reference
HAMD-17	17-Item Hamilton Depression Scale
HRQL	Health-related quality-of-life
J2R	Jump-to-reference
LOCF	Last observation carried forward
MAE	Mean absolute error
MAR	Missing at random
MCAR	missing completely at random
MI	Multiple imputation
MICE	Multiple imputation by chained equations
MMRM	Mixed model for repeated measures
MMRMD	Direct analysis without imputation
MNAR	Missing-not-at-random mechanisms
MRE	Mean relative error
PCI-active	Pain coping inventory
PFDD	Patient-focused drug development
PPMs	Pattern mixture models
PRO	Patient-reported outcome
RCT	Randomized controlled trial
RMSE	Root mean square error
SCORE-MICE	MICE at the composite score level

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12955-025-02364-0>.

Supplementary Material 1

#### Acknowledgements

None.

#### Author contributions

C.O. conceived the study; M.Y., L.Z., C.Z. and C.S. prepared and conducted the literature search. M.Y. conducted the statistical analysis and drafted the manuscript; L.Z. and C.Z. assisted with simulation study; C.O. critically revised the manuscript. All authors provided critical revision of the manuscript for important intellectual content.

#### Funding

This work was supported by Guangdong Basic and Applied Basic Research Foundation [2021A1515220065] and the National Natural Science Foundation of China [82373679].

#### Data availability

No datasets were generated or analysed during the current study.

#### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

Published online: 05 April 2025

#### References

1. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research, & U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes. 2006;4:79. <https://doi.org/10.1186/1477-7525-4-79>.
2. European Medicines Agency. (2005). Reflection Paper on the Regulatory Guidance for the Use of Health-Related Quality of Life (Hrql) Measures in the Evaluation of Medicinal Products. Retrieved July 15, 2024, from <https://zy.yaozh.com/sda/WC500003637.pdf>
3. Kyte D, Reeve BB, Efficace F, Haywood K, Mercieca-Bebber R, King MT, Norquist JM, Lenderking WR, Snyder C, Ring L, Velikova G, Calvert M. International society for quality of life research commentary on the draft European medicines agency reflection paper on the use of patient-reported outcome (PRO) measures in oncology studies. Qual Life Research: Int J Qual Life Aspects Treat Care Rehabilitation. 2016;25(2):359–62. <https://doi.org/10.1007/s11136-015-1099-z>.
4. Kluetz PG, Bhatnagar V. The FDA's patient-focused drug development initiative. Volume 19. Clinical advances in hematology & oncology: H&O; 2021. pp. 70–2. 2.
5. Rombach I, Gray AM, Jenkinson C, Murray DW, Rivero-Arias O. Multiple imputation for patient reported outcome measures in randomised controlled trials: advantages and disadvantages of imputing at the item, subscale or composite score level. BMC Med Res Methodol. 2018;18(1):87. <https://doi.org/10.1186/s12874-018-0542-6>.
6. Krepper D, Giesinger JM, Dirven L, Efficace F, Martini C, Thurner AMM, Al-Naesan I, Gross F, Sztankay MJ. Information about missing patient-reported outcome data in breast cancer trials is frequently not documented: a scoping review. J Clin Epidemiol. 2023;162:1–9. <https://doi.org/10.1016/j.jclinepi.2023.07.012>.
7. Palmer MJ, Mercieca-Bebber R, King M, Calvert M, Richardson H, Brundage M. A systematic review and development of a classification framework for factors associated with missing patient-reported outcome data. Clin Trials. 2018;15(1):95–106. <https://doi.org/10.1177/1740774517741113>.
8. Little RJ, Agostino D, Cohen R, Dickersin ML, Emerson K, Farrar SS, Frangakis JT, Hogan C, Molenberghs JW, Murphy G, Neaton SA, Rotnitzky JD, Scharfstein A, Shih D, Siegel WJ, J. P, Stern H. The prevention and treatment of missing data in clinical trials. N Engl J Med. 2012;367(14):1355–60. <https://doi.org/10.1056/NEJMs1203730>.
9. Fairclough DL, Peterson HF, Cella D, Bonomi P. Comparison of several model-based methods for analysing incomplete quality of life data in cancer clinical trials. Stat Med. 1998;17(5–7):781–96.
10. [https://doi.org/10.1002/\(sici\)1097-0258\(19980315/15\)17:5/7%3C781::aid-sim821%3E3.0.co;2-o](https://doi.org/10.1002/(sici)1097-0258(19980315/15)17:5/7%3C781::aid-sim821%3E3.0.co;2-o)
11. Mercieca-Bebber R, Palmer MJ, Brundage M, Calvert M, Stockler MR, King MT. Design, implementation and reporting strategies to reduce the instance and impact of missing patient-reported outcome (PRO) data: a systematic review. BMJ Open. 2016;6(6). <https://doi.org/10.1136/bmjopen-2015-010938>.
12. Stang A. Nonresponse research—an underdeveloped field in epidemiology. Eur J Epidemiol. 2003;18(10):929–31. <https://doi.org/10.1023/a:1025877501423>.
13. Worboys HM, Cooper NJ, Burton JO, Young HML, Waheed G, Fotheringham J, Gray LJ. Measuring quality of life in trials including patients on haemodialysis: methodological issues surrounding the use of the kidney disease quality of life questionnaire. Nephrology, dialysis, transplantation: official publication of the European Dialysis and transplant association. - Eur Ren Association. 2022;37(12):2538–54. <https://doi.org/10.1093/ndt/gfac170>.
14. Carpenter JR, Smuk M. Missing data: A statistical framework for practice. Biom J. 2021;63(5):915–47. <https://doi.org/10.1002/bimj.202000196>.
15. Kenward MG, Lesaffre E, Molenberghs G. An application of maximum likelihood and generalized estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random. Biometrics. 1994;50(4):945–53.
16. Doidge JC. Responsiveness-informed multiple imputation and inverse probability-weighting in cohort studies with missing data that are non-monotone

- or not missing at random. *Stat Methods Med Res.* 2018;27(2):352–63. <https://doi.org/10.1177/0962280216628902>.
17. Lane P. Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. *Pharm Stat.* 2008;7(2):93–106. <https://doi.org/10.1002/pst.267>.
  18. Carpenter J, Kenward MG. (2007). Missing data in randomised controlled trials: a practical guide.
  19. Fairclough DL. Design and analysis of quality of life studies in clinical trials. *Qual Life Res.* 2002;13:275–7.
  20. Machin D, Fayers PM, Fayers PM. Quality of life. Assessment, Analysis, and Interpretation; 2000.
  21. Barnes SA, Mallinckrodt CH, Lindborg SR, Carter MK. The impact of missing data and how it is handled on the rate of false-positive results in drug development. *Pharm Stat.* 2008;7(3):215–25. <https://doi.org/10.1002/pst.310>.
  22. Siddiqui O, Hung HM, O'Neill R. MMRM vs. LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets. *J Biopharm Stat.* 2009;19(2):227–46. <https://doi.org/10.1080/10543400802609797>.
  23. Fielding S, Fayers P, Ramsay CR. Analysing randomised controlled trials with missing data: choice of approach affects conclusions. *Contemp Clin Trials.* 2012;33(3):461–9. <https://doi.org/10.1016/j.cct.2011.12.002>.
  24. Marshall A, Altman DG, Royston P, Holder RL. Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. *BMC Med Res Methodol.* 2010;10:7. <https://doi.org/10.1186/1471-2288-10-7>.
  25. Fielding S, Fayers P, Ramsay C. Predicting missing quality of life data that were later recovered: an empirical comparison of approaches. *Clin Trials.* 2010;7(4):333–42. <https://doi.org/10.1177/1740774510374626>.
  26. Strong B, Fritz MC, Woodward A, Kozlowski A, Reeves MJ. Responder analysis confirms results of a stroke transitional care trial but provides more interpretable results. *J Clin Epidemiol.* 2023;156:66–75. <https://doi.org/10.1016/j.jclinepi.2023.01.009>.
  27. Gugliotta M, da Costa BR, Dabis E, Theiler R, Jüni P, Reichenbach S, Landolt H, Hasler P. Surgical versus Conservative treatment for lumbar disc herniation: a prospective cohort study. *BMJ Open.* 2016;6(12):e012938. <https://doi.org/10.1136/bmjopen-2016-012938>.
  28. Paller AS, Seyger MMB, Magariños GA, Pinter A, Cather JC, Rodríguez-Capriles C, Zhu D, Somani N, Garrelts A, Papp KA, IXORA-PEDS Investigators. Long-term efficacy and safety of up to 108 weeks of Ixekizumab in pediatric patients with moderate to severe plaque psoriasis: the IXORA-PEDS randomized clinical trial. *JAMA Dermatology.* 2022;158(5):533–41. <https://doi.org/10.1001/jamadermatol.2022.0655>.
  29. Campbell L, Ibrahim F, Barbini B, Samarawickrama A, Orkin C, Fox J, Waters L, Gilleece Y, Tariq S, Post FA. Bone mineral density, kidney function and participant-reported outcome measures in women who switch from tenofovir disoproxil emtricitabine and a nonnucleoside reverse transcriptase inhibitor to abacavir, lamivudine and dolutegravir. *HIV Med.* 2022;23(4):362–70. <https://doi.org/10.1111/hiv.13215>. & BEST Trial Team\*.
  30. Castellano D, del Muro XG, Pérez-Gracia JL, González-Larriba JL, Abrio MV, Ruiz MA, Pardo A, Guzmán C, Cerezo SD, Grande E. Patient-reported outcomes in a phase III, randomized study of Sunitinib versus interferon- $\alpha$  as first-line systemic therapy for patients with metastatic renal cell carcinoma in a European population. *Annals Oncology: Official J Eur Soc Med Oncol.* 2009;20(11):1803–12. <https://doi.org/10.1093/annonc/mdp067>.
  31. Latocha KM, Løppenthin KB, Østergaard M, Jennum PJ, Hetland ML, Røgind H, Lundbak T, Midtgaard J, Christensen R, Esbensen BA. The effect of group-based cognitive behavioural therapy for insomnia in patients with rheumatoid arthritis: a randomized controlled trial. *Rheumatology (Oxford).* 2023;62(3):1097–107. <https://doi.org/10.1093/rheumatology/keac448>.
  32. Cella D, Ivanescu C, Holmstrom S, Bui CN, Spalding J, Fizazi K. Impact of enzalutamide on quality of life in men with metastatic castration-resistant prostate cancer after chemotherapy: additional analyses from the AFFIRM randomized clinical trial. *Annals Oncology: Official J Eur Soc Med Oncol.* 2015;26(1):179–85. <https://doi.org/10.1093/annonc/mdu510>.
  33. Carpenter JR, Roger JH, Kenward MG. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. *J Biopharm Stat.* 2013;23(6):1352–71. <https://doi.org/10.1080/10543406.2013.834911>.
  34. Jin M, Fang Y. On Reference-based imputation for analysis of incomplete repeated binary endpoints. *J Biopharm Stat.* 2022;32(5):692–704. <https://doi.org/10.1080/10543406.2021.2011899>.
  35. Lipkovich I, Ratitch B, Mallinckrodt C. Causal inference and estimands in clinical trials. *Stat Biopharm Res.* 2020;12:54–67.
  36. Little R, Yau L. Intent-to-treat analysis for longitudinal studies with drop-outs. *Biometrics.* 1996;52(4):1324–33.
  37. Liu GF, Pang L. On analysis of longitudinal clinical trials with missing data using reference-based imputation. *J Biopharm Stat.* 2016;26(5):924–36. <https://doi.org/10.1080/10543406.2015.1094810>.
  38. Jin M. Imputation methods for informative censoring in survival analysis with time dependent covariates. *Contemp Clin Trials.* 2024;136:107401. <https://doi.org/10.1016/j.cct.2023.107401>.
  39. National Research Council. The prevention and treatment of missing data. Clinical trials. Panel on handling missing data in clinical trials. Committee on National statistics, division of behavioral and social sciences and education. Washington, DC: National Academies; 2010. <https://doi.org/10.17226/12955>.
  40. European Medicines Agency. (2010). Guideline on Missing Data in Confirmatory Clinical Trials. [https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-missing-data-confirmatory-clinical-trials\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-missing-data-confirmatory-clinical-trials_en.pdf)
  41. Curran D, Molenberghs G, Fayers PM, Machin D. Incomplete quality of life data in randomized trials: missing forms. *Stat Med.* 1998;17(5–7):697–709. [https://doi.org/10.1002/\(sici\)1097-0258\(19980315/15\)17:5/7%3C697::aid-sim815%3E3.0.co;2-y](https://doi.org/10.1002/(sici)1097-0258(19980315/15)17:5/7%3C697::aid-sim815%3E3.0.co;2-y).
  42. Fayers PM, Curran D, Machin D. Incomplete quality of life data in randomized trials: missing items. *Stat Med.* 1998;17(5–7):679–96. [https://doi.org/10.1002/\(sici\)1097-0258\(19980315/15\)17:5/7%3C679::aid-sim814%3E3.0.co;2-x](https://doi.org/10.1002/(sici)1097-0258(19980315/15)17:5/7%3C679::aid-sim814%3E3.0.co;2-x).
  43. Simons CL, Rivero-Arias O, Yu LM, Simon J. Multiple imputation to deal with missing EQ-5D-3L data: should we impute individual domains or the actual index? *Qual Life Research: Int J Qual Life Aspects Treat Care Rehabilitation.* 2015;24(4):805–15. <https://doi.org/10.1007/s11136-014-0837-y>.
  44. Eekhout I, de Vet HC, Twisk JW, Brand JP, de Boer MR, Heymans MW. Missing data in a multi-item instrument were best handled by multiple imputation at the item score level. *J Clin Epidemiol.* 2014;67(3):335–42. <https://doi.org/10.1016/j.jclinepi.2013.09.009>.
  45. Tseng CH, Elashoff R, Li N, Li G. *Stat Methods Med Res.* 2016;25(1):205–20. <https://doi.org/10.1177/096228021448721>. Longitudinal data analysis with non-ignorable missing data.
  46. Rombach I, Jenkinson C, Gray AM, Murray DW, Rivero-Arias O. Comparison of statistical approaches for analyzing incomplete longitudinal patient-reported outcome data in randomized controlled trials. *Patient Relat Outcome Measures.* 2018;9:197–209. <https://doi.org/10.2147/PROM.S147790>.
  47. Mallinckrodt CH, Kaiser CJ, Watkin JG, et al. Type I error rates from likelihood-based repeated measures analyses of incomplete longitudinal data. *Pharm Stat.* 2004;3(3):171–86.
  48. Mallinckrodt C, Clark W, David SR. Type I error rates from mixed effects model repeated measures versus fixed effects ANOVA with missing values imputed via last observation carried forward. *Drug Inform J.* 2001;35(4):1215–25.
  49. Little RJA. Pattern-mixture models for multivariate incomplete data. *J Am Stat Assoc.* 1993;88(421):125–34.
  50. Little RJA. A class of pattern-mixture models for normal incomplete data. *Biometrika.* 1994;81(3):471–83.
  51. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res.* 2011;20(1):40–9. <https://doi.org/10.1002/mpr.329>.
  52. Lee KJ, Simpson JA. Introduction to multiple imputation for dealing with missing data. *Respirol (Carlton Vic).* 2014;19(2):162–7. <https://doi.org/10.1111/resp.12226>.
  53. Siddiqui O, Ali MW. A comparison of the random-effects pattern mixture model with last-observation-carried-forward (LOCF) analysis in longitudinal clinical trials with dropouts. *J Biopharm Stat.* 1998;8(4):545–63. <https://doi.org/10.1080/10543409808835259>.
  54. Rubin DB. (1975). inference and missing data. *Psychometrika*, 1975, 19.
  55. Molenberghs G, Kenward MG. (2007). Missing Data in Clinical Studies.
  56. Sijtsma K, van der Ark LA. Investigation and treatment of missing item scores in test and questionnaire data. *Multivar Behav Res.* 2003;38(4):505–28. [https://doi.org/10.1207/s15327906mbr3804\\_4](https://doi.org/10.1207/s15327906mbr3804_4).
  57. 50Hardouin JB, Conroy R, Sébille V. Imputation by the mean score should be avoided when validating a patient reported outcomes questionnaire by a Rasch model in presence of informative missing data. *BMC Med Res Methodol.* 2011;11:105. <https://doi.org/10.1186/1471-2288-11-105>.
  58. Bono C, Ried LD, Kimberlin C, Vogel B. Res social administrative pharmacy: RSAP. 2007;3(1):1–27. <https://doi.org/10.1016/j.sapharm.2006.04.001>. Missing data on the Center for Epidemiologic Studies Depression Scale: a comparison of 4 imputation techniques.
  59. Liu SH, Chrysanthopoulou SA, Chang Q, Hunnicutt JN, Lapane KL. Missing data in marginal structural models: A plasmode simulation study



- comparing multiple imputation and inverse probability weighting. *Med Care*. 2019;57(3):237–43. <https://doi.org/10.1097/MLR.0000000000001063>.
60. Shrive FM, Stuart H, Quan H, Ghali WA. Dealing with missing data in a multi-question depression scale: a comparison of imputation methods. *BMC Med Res Methodol*. 2006;6:57. <https://doi.org/10.1186/1471-2288-6-57>.
  61. Wang S, Schwartz PF, Mancuso JP. Comprehensive implementations of multiple imputation using retrieved dropouts for continuous endpoints. *BMC Med Res Methodol*. 2025;25(1):47. <https://doi.org/10.1186/s12874-025-0249-4-5>.
  62. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med*. 2019;38(11):2074–102. <https://doi.org/10.1002/sim.8086>.
  63. R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
  64. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. *Stat Methods Med Res*. 2014;23(5):440–59. <https://doi.org/10.1177/0962280213476378>.
  65. Fielding S, Maclennan G, Cook JA, Ramsay CR. A review of RCTs in four medical journals to assess the use of imputation to overcome missing data in quality of life outcomes. *Trials*. 2008;9:51. <https://doi.org/10.1186/1745-6215-9-51>.
  66. Rombach I, Rivero-Arias O, Gray AM, Jenkinson C, Burke Ó. The current practice of handling and reporting missing outcome data in eight widely used proms in RCT publications: a review of the current literature. *Qual Life Research: Int J Qual Life Aspects Treat Care Rehabilitation*. 2016;25(7):1613–23. <https://doi.org/10.1007/s11136-015-1206-1>.
  67. Goshio M, Maruo K. An application of the mixed-effects model and pattern mixture model to treatment groups with differential missingness suspected not-missing-at-random. *Pharm Stat*. 2021;20(1):93–108. <https://doi.org/10.1002/pst.2058>.

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.