

Multiple Imputation Methods for Handling Missing Data under Treatment Policy Strategy: A Scoping Review

Mingxin Jiang

`mingxin.jiang23@imperial.ac.uk`

Imperial College London

Leila Janani

Imperial College London

Suzie Cro

Imperial College London

Research Article

Keywords: Multiple Imputation, Treatment Policy Strategy, Missing Data, Randomised Clinical Trial, Estimand

Posted Date: September 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10908258/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

The ICH E9 (R1) addendum introduced the estimand framework for randomised controlled trials. Under a treatment policy strategy, treatment effects are evaluated regardless of intercurrent events; however, post-intercurrent-event outcomes are often missing, and standard multiple imputation (MI) may not reflect the required assumptions. This scoping review aimed to identify and summarise MI methods for handling missing data when targeting a treatment policy strategy.

Methods

We conducted a scoping review of MEDLINE and Embase from inception to 8 January 2026. An iterative search strategy comprised an initial search for MI under the treatment policy strategy followed by targeted searches for four method categories identified during the review: reference-based (RB), retrieved dropout (RD), return-to-baseline (RTB), and washout methods.

Results

The search identified 447 records, of which 52 articles were included after screening and eligibility assessment. Of these, 26 (50.0%) proposed or extended MI methods, 17 (32.7%) compared MI methods, six (11.5%) addressed uncongeniality, two (3.8%) were tutorials or overviews, and two (3.8%) addressed other topics; one article contributed to both the development and comparison categories. Proposed or extended methods included RB, RD, RTB, washout, and hybrid approaches applicable to a treatment policy strategy. We further classified approaches by outcome type and imputation framework, including RB assumptions (e.g., jump to reference), RD imputation structures (e.g., Current models), and the statistical models used for implementation.

Conclusions

A range of MI approaches can support estimation under a treatment policy strategy across different outcome types, but they rely on distinct assumptions and have different limitations. This review maps these approaches and their implementation to support principled method selection and to highlight areas where further methodological development and comparative evaluation are needed.

Background

The ICH E9 (R1) addendum marked a major shift in clinical trial analysis by formally introducing the estimand framework, including explicit consideration of intercurrent events (ICEs) [1]. ICEs are defined as events that affect either the interpretation or existence of the outcome data following the start of

treatment [1]. To handle these events, the addendum outlines five distinct strategies, each corresponding to a different question of interest. The treatment policy strategy addresses the effect of assigning treatment regardless of the occurrence of intercurrent events, such as treatment discontinuation or non-adherence [2]. Under the treatment-policy strategy, all outcomes are analysed as observed, irrespective of whether they are collected before or after the occurrence of ICEs. This requires collecting outcome data following ICEs such as treatment withdrawal; however, such data are often missing after treatment withdrawal, which complicates the analysis. The analysis of clinical trial data with missing observations is always based on untestable assumptions regarding the missing data mechanism. If the assumptions do not adequately represent the unobserved data, missing data can introduce significant bias and undermine the validity of trial results. Therefore, efficient and appropriate methods to handle missing data when treatment policy strategies are employed are required to alleviate this issue.

Multiple Imputation (MI) has been developed as a principled method for imputing missing outcome data under the assumption that the missing data follows the distribution of observed outcome data conditional on observed variables (known as missing at random (MAR)). It generates multiple copies of full data sets where the unobserved data are imputed by values sampled from a posterior distribution given the observed data. Each imputed data set is analysed using the pre-specified substantive analysis model of interest that would have been used in the absence of missing data, and the resulting estimates are combined for final inference using Rubin's rules [3].

However, standard MI under MAR implicitly assumes that participants with missing outcomes continue to behave similarly to participants in the same treatment arm with the same observed variables included in the imputation model after dropout. This assumption may not align with the treatment policy strategy discussed previously. In particular, participants who discontinue treatment may no longer follow the same outcome trajectory as those remaining on treatment. Therefore, when data are missing following ICEs, alternative imputation approaches that depart from the standard MAR MI may be warranted. Wang et al. [4] outlined four broad categories of MI methods applicable under the treatment-policy strategy: Retrieved Dropout (RD) imputation, Reference-Based (RB) imputation, washout imputation, and Return-to-Baseline (RTB) imputation.

RD MI is an effective imputation approach designed for handling missing data under a treatment policy strategy. It imputes off-treatment data based solely on observed off-treatment data [5]. Alternatively, RB MI could be used under this scenario, which assumes the patients' off-treatment outcome data behaves similarly to the data of those observed in a specified reference group, most commonly the placebo or active control arm [6]. Four common assumptions for RB MI are copy reference (CR), jump to reference (J2R), copy increments in reference (CIR), and last mean carried forward (LMCF). The third MI approach discussed by Wang et al. [4], washout imputation, is conceptually similar to RB MI in that it assumes the trajectory for missing outcomes is the same as for completers in the placebo/control arm. However, unlike RB MI, which imputes all missing time points, washout imputation imputes only the primary endpoint at the final visit. Under this approach, the experimental treatment effect is washed out among patients who withdraw, while only the placebo effect is retained [4]. Finally, RTB multiple imputation

discussed by Wang et al. [4] assumes that following treatment withdrawal, patient's missing outcomes are imputed using their baseline value as the estimated mean with added random variation [7].

Although Wang et al. [4] provided a methodological overview and comparison of four broad imputation approaches applicable to the treatment policy strategy, additional methodological variants based on alternative underlying statistical models are available. For example, the RD MI framework encompasses a wide range of model variants [5, 8–10]. Only some of which were considered by Wang et al. [4]. In addition, their overview focused primarily on continuous outcomes, whereas several methods can also accommodate other outcome types. These methodological variants and extensions warrant further consideration within a comprehensive review of the available MI methods. This review could guide researchers in selecting the most appropriate approach for their context and to identify potential gaps in the literature, such as outcome types for which MI methods under the treatment policy strategy remain underdeveloped. To address this need, this scoping review of the methodological literature aims to identify and summarise MI methods for handling missing data in clinical trials targeting a treatment policy strategy.

Methods

Search Strategy

Martin et al. [11] provided a practical framework for conducting methodology scoping reviews. Consistent with the recommendations of Martin et al. [11], we adopted an iterative search strategy because it was difficult to define search terminology a priori in this emerging research area. Searches of the MEDLINE (Ovid) and Embase (Ovid) databases were conducted on January 8, 2026, covering all records available from database inception to that date.

First, we included terms related to the treatment policy strategy, including “treatment policy” or “de facto”, as well as terms describing off-treatment, post-deviation, post-intercurrent event, and post-discontinuation data or outcomes. The truncated term “imput*” was used to capture imputation-related records. Furthermore, additional method-related terminology, including terms for the four broad categories of MI methods (RB MI, RD MI, washout MI, and RTB MI), was identified from Wang et al. [4] and incorporated into the final search strategy as part of the iterative search process, consistent with the recommendations of Martin et al. [11]. Synonymous or related terms within each concept were combined using OR, and the resulting concept groups were combined using AND. Searches were performed using the Ovid “.mp.” field, which searches multiple fields, including title, abstract, heading words, and keyword heading words. The full search strategy, including method-related terminology variants, is provided in the Additional file 2 (Table S.1).

Search Selection

The review aimed to identify methodological studies and reviews on MI methods relevant to the treatment policy strategy. The search covered a broad range of MI approaches, including RB, RD, washout, RTB, and related methods. Eligible studies included methodological papers or reviews investigating imputation methods under a treatment policy strategy that included MI approaches. Only full-text articles published in English were considered. Full eligibility criteria are provided in Additional file 2.

All records identified through the database searches were imported into Covidence, with duplicates removed using Covidence and through subsequent manual checking by MJ. Titles and abstracts were then screened for eligibility by a single reviewer (MJ). Full-text articles were subsequently assessed for eligibility by one reviewer (MJ). Where eligibility was uncertain, inclusion was discussed with SC and LJ to reach a final decision.

Data Extraction and Statistical Methods of Review

A data extraction form was developed and implemented within Covidence to systematically collect bibliographic information, methodological study characteristics, details of the imputation methods investigated, treatment policy and estimand-related information, and aspects related to reproducibility. Further details of the data extraction sheet are provided in Additional file 3. Data extraction was performed by a single reviewer (MJ), with uncertainties resolved through discussion with a second reviewer (SC or LJ) where necessary.

After data extraction, the extracted data were organised and managed in Microsoft Excel for further analysis. Each identified MI method was reviewed in detail to understand its overall characteristics, including the underlying statistical model used for imputation. Methods were categorised using the four main broad types of MI methods previously discussed and any combinations of these, namely RB, RD, RTB and washout. The extracted information was summarised descriptively using frequencies and percentages and presented in tabular form. Due to the methodological nature of the included studies, no formal quantitative synthesis or meta-analysis was performed.

Results

The search yielded a total of 447 articles. After duplicates were removed, 315 articles were screened based on their titles and abstracts. Of these, 241 articles were excluded because they did not meet the eligibility criteria. The main reasons for exclusion were that the articles reported applied clinical trial results without a methodological focus, were commentary papers, conference abstracts, cost-effectiveness studies, or were unrelated to the review topic. The remaining 74 articles underwent full-text assessment for eligibility. Following full-text review, 22 articles were excluded because they did not meet the inclusion criteria. Ultimately, 52 studies met the eligibility criteria and were included in the review for data extraction, as illustrated in the PRISMA flow diagram (Fig. 1).

Characteristics of the Included Papers

The key characteristics of the included papers are shown in Table 1. More details can be found in Additional file 2 (Table S.2). Of the 52 included studies, the majority were methodological study papers, and 1 was a methodological scoping review. Among the included articles, more papers were published after 2019, and these often referred to the ICH E9 (R1) Addendum guideline and explicitly mentioned the concept of the treatment policy strategy (Fig. 2). Prior to 2019, similar concepts were often described using alternative terms, including de facto estimand, effectiveness estimand, and ITT estimand. The majority of articles originated from the United States of America (USA) (27; 51.9%), followed by the United Kingdom (UK) (15; 28.8%).

Table 1
Characteristics of included articles (N = 52)

Category	N	%
Journal		
Pharmaceutical Statistics	14	26.9
Statistics in Biopharmaceutical Research	8	15.4
Statistics in Medicine	6	11.5
Journal of Biopharmaceutical Statistics	6	11.5
Contemporary Clinical Trial	5	9.6
Statistical Methods in Medical Research	2	3.8
BMC Medical Research Methodology	2	3.8
Biometrics	2	3.8
Other	7	13.5
Country¹		
United States	27	51.9
United Kingdom	15	28.8
China	3	5.8
Switzerland	3	5.8
Australia	2	3.8
Spain	2	3.8
Type of Paper		
Methodology Paper	51	98.1
Review Paper	1	1.9
Imputation Method – Treatment Policy Alignment²		
RB	40	76.9
RD	11	21.2
RTB	5	9.6
Washout	3	5.8
<i>Notes: ¹ Based on the first author's first affiliation. ² Multiple elements could be included in a single paper.</i>		

Category	N	%
Journal		
Combination Imputation Methods		
RD + RB		
– RD RBI-C	1	1.9
– RD_p	1	1.9
RD + RTB		
– RD-RTB	1	1.9
Scoping Review included RD, RB and RTB	1	1.9
Types of Methodological Study Elements²		
Case Study	40	76.9
Simulation Study	37	71.2
Analytical Study	1	1.9
Tutorial / Overview	1	1.9
Scoping Review	1	1.9
Code Available for Imputation		
Yes	32	61.5
– SAS	21	40.4
– R	8	15.4
– Stata	2	3.8
– Stata and R	1	1.9
No	18	34.6
– SAS	8	15.4
– R	1	1.9
– Stata	1	1.9
– Not Reported	8	15.4
Without code using	2	3.8
<i>Notes: ¹ Based on the first author's first affiliation. ² Multiple elements could be included in a single paper.</i>		

Category	N	%
Journal		
– Tutorial / Overview	1	1.9
– Scoping Review	1	1.9
Code Available for Simulation		
Yes	12	23.1
– R	8	15.4
– SAS	3	5.9
– R and SAS	1	1.9
No	24	46.2
Without simulation study	16	30.8
<i>Notes: ¹ Based on the first author's first affiliation. ² Multiple elements could be included in a single paper.</i>		

[Insert Table 1 here]

With respect to treatment policy-aligned imputation approaches, four broad categories of methods were identified: Reference-Based (RB) imputation (40; 76.9%), Retrieve Dropout (RD) imputation (11; 21.2%), Return-to-Baseline (RTB) imputation (5; 9.6%), and washout imputation (5; 5.8%). In addition, three hybrid imputation methods were identified. Two of these hybrid methods, Retrieved Dropout Reference-Based Centred Multiple Imputation (RD RBI-C) and RD_p (referred by Yuan et al. [12]), combined features of RD and RB MI. The third method, RD-RTB, combined features of RD and RTB MI. These methods are described in greater detail in the “Development of New Multiple Imputation Methods” Section.

In terms of methodological study elements included in the paper, the papers employed a range of approaches. Case studies (40; 76.9%) and simulation studies (37; 71.2%) were the most frequently reported. In contrast, analytical studies, tutorials/overviews, and scoping reviews were each represented by only one article (1.9%).

Regarding software implementation, 42 articles (80.8%) reported the software used for imputation and analysis, of which 32 (61.5%) provided accompanying code. SAS was the most frequently used software for imputation (21; 40.4%), followed by R (8; 15.4%) and Stata (2; 3.8%). One article (1.9%) reported implementations in both R and Stata, where each software corresponded to a different imputation model: the multivariate normal (MVN) model with a rounding algorithm implemented in Stata, whereas the latent MVN algorithm implemented in R. Among simulation studies, only 12 articles (23.1%) reported the code used for generating simulation data, with R being the most frequently adopted software.

The included papers were classified into four main primary research directions: (1) the development of new imputation methods, (2) comparative evaluation of imputation methods, (3) solving the uncongeniality issue under MI, and (4) tutorials or overviews (Table 2). Additionally, two papers that did not fall within any of these categories were grouped into an “Other” category, encompassing topics such as sensitivity analysis properties and scoping reviews.

Table 2
Classification of included studies according to research direction

Category ¹	N	%
Development of New Multiple Imputation Methods	26	50.0
Comparative Evaluation of Imputation Methods	17	32.7
Solving Uncongeniality Issue under Multiple Imputation	6	11.5
Tutorials / Overviews	2	3.8
Other ²	2	3.8
<i>Notes: ¹ One study was classified in both categories, as it proposed a new imputation method while also conducting a comparative evaluation. ² Includes studies addressing sensitivity analysis properties and scoping reviews.</i>		

Development of New Multiple Imputation Methods

Of the 52 articles identified in the scoping review, 26 (50%) focused on the development of new imputation methods (Table 2). Most of these studies extended existing imputation frameworks rather than proposing entirely new approaches. Specifically, 15 articles proposed or extended RB MI methods, 7 proposed or extended RD MI methods, and one extended the RTB framework. In addition, three studies developed hybrid MI methods that combined features of previous approaches (Fig. 3).

Among the papers developing MI methods, 12 focused on methods for handling missingness in longitudinal continuous outcomes under the treatment policy strategy (Table 3). The MVN model was the predominant choice for imputing continuous outcomes, being used in 6 of the 12 studies (50%). For RB MI methods, it could be implemented using either the MVN model or sequential linear regression. Notably, one study extended this framework to accommodate nonnormally distributed continuous outcomes by employing sequential weighted robust regression instead of standard normality-based models [13]. Additionally, White et al. [14] extended the RB method using the MVN model under the causal framework. For RD MI methods, it could be classified into two categories according to the imputation strategy: Last-visit imputation models and Repeated-measure imputation models. Linear regression was typically used for last-visit-based imputation, whereas repeated-measure imputation was implemented using either the MVN model or sequential linear regression. One study proposed two alternative implementations for imputation: a two-step sequential linear regression approach and a one-step MVN model implemented via MCMC for monotone and non-monotone missing data, respectively

[8]. For the RTB method, the imputation model was based on an MVN. In addition, three hybrid imputation approaches for continuous outcomes have been proposed. RD RBI-C was based on an MVN model, RD_p was based on sequential linear regression, and RD-RTB was based on an MVN (Table 3).

Table 3
Summary of multiple imputation for continuous outcome

Imputation Methods/ Categorisation(s)	N	Studies	Statistical Models
RD			
Last-visit imputation model:			
Split-dataset	1	Wang (2022) [5]	Linear regression
Repeated-measure imputation model:			
Split-dataset	1	Wang (2023) [8]	Two-step: sequential linear regression; one-step: MVN
Current & Historic	2	Drury (2024) [10]	Sequential linear regression
		Noci (2023) [15]	MVN
RB			
J2R	1	Liu (2024) [13]	Sequential weighted robust regressions for non-normal assumption
J2R & CR	1	Fang (2022) [16]	Sequential linear regression
J2R & CR & CIR	2	Tang (2017) [17]	MVN
		White (2020) [14]	MVN under causal framework
J2R & CR & CIR & LMCF	1	Carpenter (2013) [6]	MVN
RD + RB			
– RD RBI-C			
RB: J2R & CIR & LMCF; RD: Current & Historic	1	Cro (2024) [18]	MVN
– RD_p			
RB: NA; RD: Current	1	Yuan (2025) [12]	Sequential linear regression
RTB	1	Zhang (2022) [7]	Normal distribution using baseline mean
<i>Note: Split-dataset: The dataset is decomposed into three subsets: subjects with missing values at the primary visit, off-treatment completers, and on-treatment completers.</i>			

Imputation Methods/ Categorisation(s)	N	Studies	Statistical Models
RD			
RD + RTB			
– RD-RTB			
Not Specific	1	Li (2025) [19]	MVN
<i>Note: Split-dataset: The dataset is decomposed into three subsets: subjects with missing values at the primary visit, off-treatment completers, and on-treatment completers.</i>			

Table 3 also further shows different imputation model categorisations for RD MI and RB MI under continuous outcomes. The RD methods share the same underlying concept of imputing missing data based on observed off-treatment data; however, each identified variant of the method differs in the observed data used and in how these data are incorporated into the imputation model. As mentioned above, RD MI methods can be broadly categorised into last-visit and repeated-measures imputation models based on what data is imputed. Under the last-visit imputation model, the imputation model uses the baseline and the last on-treatment visit to impute the primary visit data [5]. It proposed a split-dataset approach in which the full dataset is decomposed into three subsets: the complete on-treatment data, the observed off-treatment data (referred to as RD data in their paper), and the missing data. The imputation is then performed solely for the missing data subset, informed by the observed off-treatment data. Moving beyond the last-visit model, the repeated-measure model is more widely utilised among the proposed methods. One study similarly adopted a split dataset approach [8], while others employed the Current and Historic model framework [10, 15]. These two models differ in their stratification criteria for imputation. Current RD stratifies patients according to treatment group and their Current on-/off-treatment status (a binary indicator at the imputation visit), whereas Historic RD incorporates additional historical treatment discontinuation information, such as the treatment discontinuation pattern up to the imputation visit [10] or the time since treatment discontinuation [15].

While RD MI relies solely on observed off-treatment data, the RB method is an imputation approach that borrows information from a pre-defined “reference” group based on specified assumptions for imputation. Four main models that make different assumptions for missing data have been proposed under this framework: CR, J2R, CIR and LMCF. Under the CR assumption, off-treatment outcomes are assumed to follow the reference group trajectory throughout the entire study period. The J2R assumption, on the other hand, retains the patient's own treatment arm trajectory up to the point of deviation, after which patients are assumed to switch immediately to the reference-group trajectory, implying a sudden loss of efficacy. Under the CIR assumption, the treatment benefit gained before deviation is preserved, but subsequent changes in the outcome response are assumed to follow the same rate of change as in the reference group, representing a gradual loss of efficacy over time. Finally, the LMCF assumption carries forward the last observed mean from the patient's own randomised treatment effect at the point of deviation to all subsequent visits, suggesting that the treatment benefit is

maintained after discontinuation. Notably, LMCF has not been further developed beyond Carpenter et al. [6], while J2R remains the most commonly extended assumption model by others.

In addition to the standalone RD and RB approaches, hybrid models have been proposed that integrate concepts from both of these frameworks. There were two hybrid imputation approaches that combined elements of RD and RB MI. Cro et al. [18] extended the core RB framework by introducing additional parameters that make the core RB model equivalent to an RD model, resulting in the RD RBI-C method. Under this framework, Cro et al. [18] discussed how the core RB model component could be specified under J2R, CIR, or LMCF assumptions, while the RD component can be implemented under either the Historic or Current model. As noted by Cro et al. [18], under CIR, the assumed effect of treatment deviation on subsequent outcomes depends on the visit at which the deviation occurs. CIR is therefore compatible with the Historic model rather than the Current model, and the combination of CIR with the Current model was not considered. Yuan et al. [12] proposed another hybrid approach also based on RD and RB, called RD_p. This method extends the On/Off-Intercepts model with the Common Slopes Using Residuals (OICS-R) model proposed by [10], which is based on the Current RD model, by introducing a prespecified threshold for the proportion of observed off-treatment data in this method. When the proportion of observed off-treatment data falls below this threshold, additional participants are randomly selected from the completers in the control group to supplement the off-treatment data, without the need to explicitly specify an RB assumption.

Additionally, a third hybrid approach combining concepts from both RD and RTB frameworks has also been proposed. Li et al. [19] uses RTB imputation as an initial imputation model and subsequently modifies the parameters of this RTB model through an individual-specific calibration term based on baseline value and exposure time. The calibration term is further scaled according to the relative difference in mean absolute change from baseline to the final visit between off-treatment and on-treatment subjects, with the off-treatment mean estimated under the RD method. In addition, the sign of ρ is used to determine the direction of the calibration, reflecting whether the efficacy of off-treatment subjects is expected to be better or worse than that of on-treatment subjects. However, the imputation model underlying the RD approach is not explicitly described; therefore, the corresponding imputation model was recorded as not specific. For further details on this method, see Li et al. [19].

The proposed MI methods identified also include those applicable to other outcome types, including binary, categorical, ordinal, recurrent-event, and time-to-event data. The underlying statistical imputation models varied according to the outcome type, as summarised in Table 4.

Table 4
Summary of imputation models by outcome types

Outcome Types / Models	RB	RD
Binary outcomes (6)		
Multivariate probit model	Lu (2020) [20]	
MVN with adaptive rounding algorithm	Rehal (2023) [21]	
Multivariate probit model; MVN with adaptive rounding algorithm	Cro (2025) [22]	
Logistic regression	Gao (2017) [23]	
Sequential logistic regression	Fang (2022) [16]	
Sequential logistic model; Multivariate probit model	Tang (2018) [24]	
Recurrent-event outcomes (4)		
Bayesian piecewise exponential proportional intensity model with frailty	Gao (2017) [25]	
Negative binomial model	Akacha (2016) [26]	Roger (2019) [27]
Negative multinomial distribution	Keene (2014) [28]	
Time-to-event outcomes (4)		
Cox regression model	Jin (2024) [29]	
Piecewise exponential model		He (2023) [30]
Piecewise hazard model		Hartley (2022) [31]
Weibull proportional hazards model	Atkinson (2019) [32]	
Ordinal outcomes (1)		
Sequential logistic model; multivariate probit model	Tang (2018) [24]	
<i>Notes: Numbers in parentheses denote the number of studies. Multiple outcome types may be considered within a single study.</i>		

For binary outcomes, all identified studies focused on RB methods. Sequential logistic regression was the most commonly adopted approach for specifying the imputation model. Approaches based on MVN assumptions were also applied, including latent variable formulations such as multivariate probit models, as well as methods that impute discrete variables as continuous and then use adaptive rounding to recover discrete values.

For recurrent-event outcomes, Bayesian piecewise exponential proportional intensity models with frailty and negative multinomial distributions were used exclusively in RB MI methods, whereas negative binomial models were employed in both RB and RD MI methods.

For time-to-event outcomes, RB MI methods were commonly implemented using Cox regression and Weibull proportional hazards models, whereas RD MI methods were based on piecewise exponential and piecewise hazard models.

Addressing Issues of Uncongeniality

Uncongeniality occurs when the structure or form of the analysis model applied post multiple-imputation does not correspond to that of the imputation model. This can present a challenge for variance estimation following MI. This issue has received particular attention in the context of RB MI. Six of the included articles discussed this issue of uncongeniality. Specifically, the double use of reference information within RB MI (once for the actual reference patients and secondly for the imputation of active patients) results in an imposed artificial correlation between outcomes of patients in opposing treatment arms. This reduces the empirical frequentist variance of the treatment effect with greater amounts of missing data. Rubin's variance estimator applied post RB MI circumnavigates this artificially imposed correlation by drawing independent imputations within the active arm based on reference arm data, and so consequently doesn't match the empirical variance[33, 34]. Rubin's variance estimate alternatively increases appealingly as the amount of data increases. This issue was discussed in 6 (11.5%) of the reviewed studies.

To address what variance is most appropriate, Cro et al. [35] argued that the appropriate variance should be information-anchored, which means the proportional increase in variance of the RB MI estimator due to missing data should remain approximately the same as the proportional increase in variance due to missing data that would be expected under the MAR assumption [33]. On this basis, they recommended the use of Rubin's variance estimator, which was shown to provide information-anchored inference.

Despite this, some other researchers (N = 5) derived alternative estimators that aligned with the empirical variance instead of relying on Rubin's MI variance estimation. The most commonly employed alternative approach was the bootstrap variance estimator, which was extended across different outcome types [36–38].

Comparative Evaluation of Imputation Methods

We also included the studies that considered the MI applicable under the treatment policy strategy as part of the comparator in the study. 17 (32.7%) of the reviewed articles are considered as comparative studies.

Among the 17 comparative studies, seven primarily aimed to compare and evaluate different imputation methods. Five of these seven studies focused on continuous outcomes, whereas two examined time-to-

event outcomes. All seven evaluated RB methods, and three additionally included RD approaches. A summary of the imputation methods evaluated in these studies is provided in Table 5.

Table 5
Summary of comparative evaluation studies

Study ID	Imputation methods under treatment policy strategy	Other imputation methods
Continuous outcomes		
Cai 2020 [39]	RB: J2R	MAR; Complete-Case Analysis (CCA); Last Observation Carried Forward (LOCF); Baseline Observation Carried Forward (BOCF); Trimmed Mean (TM)
Noci 2023 [15]	RB: CIR; RD: time-varying MAR (TV1-MAR, TV2-MAR)	MAR
Wang 2023 [4]	RD; RTB; Washout; RB: J2R & CR	
Yan 2025 [40]	RB: J2R, CR, CIR	LOCF; Mixed Model for Repeated Measures (MMRM); Multiple Imputation by Chained Equations (MICE); Mixed Model for Repeated Measures directly at the composite score level (MMRMD); Multiple Imputation by Chained Equations at the composite score level (SCORE-MICE)
Medcalf 2025 [41]	RD; RB: J2R	CCA; MMRM; MI; TM
Time-to-event outcomes		
Lu 2015 [42]	RB: CR	Delta-adjusted
Floden 2025 [43]	RB with reference as control arm (hypothetical); observed post-progression data (treatment policy)	

Although Noci et al. [15] was classified as a comparative study based on its primary objective, it also introduced an MVN-based RD MI method, representing a methodological contribution beyond comparative evaluation. This approach was subsequently included as one of the comparator methods in Medcalf et al. [41], where both RD and RB under the J2R assumption demonstrated the best performance under missing not at random (MNAR) assumptions. In addition, Wang et al. [4] provided a comparison of imputation methods applicable under the treatment policy strategy and offered a detailed description of several approaches, including RD, RTB, Washout, and RB methods. Further, one study compared different estimand strategies and the corresponding imputation methods used to handle missing data [43]. In this study, RB methods were used as sensitivity analyses for the MNAR assumption

under both the hypothetical and treatment policy strategies. Under the hypothetical strategy, the control arm served as the reference group for the RB methods. In contrast, under the treatment policy strategy, observed post-progression outcomes and adverse events were used as the reference data for the RB methods.

The remaining 10 included comparative studies primarily evaluated non-MI methods, such as likelihood-based, distribution-based, or analytical approaches, rather than developing methods within the MI framework. Nevertheless, these studies were included as comparative studies because they evaluated MI methods applicable under the treatment policy strategy alongside non-MI approaches. Of these 10 papers, seven focused on RB imputation and one on RD imputation. Of the remaining two, one developed distributional imputation approaches for RB, washout, and RTB, while the other developed a likelihood-based imputation framework covering the same three methods as well as RD (Table 6).

As discussed in the previous section on Addressing Issues of Uncongeniality, RB MI is associated with uncongeniality issues. Consequently, an alternative pathway is to implement RB methods without relying on the MI framework. Among the 9 papers that included non-MI implementations of RB methods, 6 considered uncongeniality as one of the motivations for avoiding the MI framework.

In addition, Bell et al. [9] proposed a likelihood-based framework for implementing RD and compared the proposed RD MMRM approach with RD MI under the Historic and Current models described by Drury et al. [10], as well as with RB MI methods. Focusing on the RD model, MMRM and MI yielded nearly identical point estimates, although the model-based standard errors from MMRM were less accurate. Furthermore, compared with RD methods, RB MI under J2R achieved substantially higher statistical power when post-ICE data were limited.

Table 6

Summary of multiple imputation methods in non-multiple imputation methodological studies

Study ID	MI methods under treatment policy strategy	Non-MI methods
Continuous outcomes		
Lu 2014 [44]	RB MI (placebo-based)	RB (placebo-based): 1) Posterior simulation 2) Maximum likelihood estimator
Tang 2015 [45]	RB MI (control-based)	RB (control-based): Likelihood-based imputation
Mehrotra 2017 [46]	RB MI (J2R)	RB (control-based): Direct estimation for control-based mean imputation
Wolbers 2022 [34]	RB MI (J2R, CR, CIR)	RB (J2R, CR, CIR): Conditional mean imputation
García-Hernández 2023 [47]	RB MI (J2R, CR, CIR, Average LMCF, Own LMCF)	RB (J2R, CR, CIR, Average LMCF, Own LMCF): Analytical approach
Bell 2025 [9]	RB MI (J2R, CR, CIR), RD MI	RD: Likelihood-based imputation
Lou 2025 [48]	RB MI (J2R), RTB MI, Washout MI, RD MI	RB (J2R), RTB, Washout, RD: Direct likelihood estimation approach
Time-to-event outcomes		
Liu 2023 [49]	RB MI (J2R), RTB MI, Washout MI	RB (J2R), RTB, Washout: Distributional imputation
García-Hernández 2024 [50]	RB MI (J2R, CIR)	RB (J2R, CIR): Maximum likelihood approach
Recurrent-events outcomes		
Fairfax 2024 [51]	RB MI (J2R, CR)	RB (J2R, CR): Distributional imputation

Tutorials/ Overview articles and Other Types of Articles

Two (3.9%) papers were classified as Tutorials/Overviews. Kenward [52] presented a general introduction to control-based imputation methods, whereas Cro et al. [53] presented sensitivity analyses for continuous outcomes and described RB MI as one of the available approaches for addressing missing data under a treatment policy strategy.

There were two other types of included studies. One was a scoping review by Medcalf et al. [54] which provided a comprehensive methodological mapping of strategies for handling various types of missing data in randomised controlled trials. The scoping review systematically mapped methods for handling missing outcome data in randomised controlled trials, including approaches developed under the MAR, MNAR, and hybrid assumptions. It summarised the strengths, limitations, and recommended applications of these methods and developed a four-step guide to assist trialists in selecting appropriate approaches for handling missing outcome data. In addition, the review explicitly reported estimand-related information in the characteristics table, including the classification of studies according to the ITT/de facto/treatment policy estimand. Notably, the authors highlighted that the standard MI method under the MAR assumption is not always appropriate for the treatment policy strategy. Another paper from Liu and Pang [55] explored the properties of control-based imputation for sensitivity analyses and introduced a new sensitivity measure based on a weighted combination of MAR and J2R assumptions.

Discussion

Summary of Principal Findings

This review identified 52 articles related to imputation methods for handling missing data in clinical trials under the treatment-policy strategy. Although publications increased following the release of the ICH E9(R1) Addendum, our findings suggest that the core frameworks for imputing data under a treatment policy strategy were already present in the literature before the formal introduction of the estimand framework. Prior to 2019, similar concepts were often described using alternative terms, such as de facto estimand, effectiveness estimand, and ITT estimand. The publication of the ICH E9(R1) Addendum appears to have facilitated the standardisation of terminology, leading to more frequent and explicit use of methods for implementing a treatment policy strategy in subsequent publications. Nevertheless, a small number of recent studies continue to employ earlier terminology, indicating that the estimand framework has not yet been fully adopted. Beyond changes in terminology, the reviewed studies showed continued methodological development in addressing missing data under a treatment policy strategy. Researchers proposed new imputation methods and extended existing approaches. These methodological extensions occurred in several directions. Some studies retained the same underlying imputation concept, such as RB and RD, while adopting alternative statistical modelling frameworks, whereas others extended existing methods to accommodate a broader range of outcome types, including binary, ordinal, recurrent-event, and time-to-event endpoints. This progression reflects efforts to accommodate the diversity and complexity of outcomes encountered in real-world clinical trials. Furthermore, several studies combined elements from different existing imputation methods,

resulting in more flexible hybrid approaches that aimed to overcome the limitations of any single method.

When implementing RB MI under the treatment-policy strategy, as shown by the identified methodological literature, the issue of uncongeniality should be carefully considered. While Rubin's rules remain widely used post RB MI, several studies have noted that variance estimation under RB MI is complicated by the uncongeniality between the imputation and analysis models [35–38]. Although many studies have recognised the variance estimation issue in RB MI, there is still no universally accepted best approach. Cro et al. [35] proposed that Rubin's variance estimate is appropriate for use as it provides information-anchored inference. Whereas Yang et al. [36], Quan et al. [37], Bartlett [38] suggest instead of using Rubin's variance estimator to use a bootstrap variance estimator. Therefore, when applying RB MI, careful consideration should be given to the choice of variance estimation method.

In light of these methodological and implementation challenges, the availability of code becomes particularly important. More than half of the reviewed articles provided code or a code template for their imputation methods. Providing code in methodological papers can help readers better understand the practical implementation of the method and promote its broader use. In contrast, simulation code was much less frequently provided, making it more difficult for readers to fully understand the data-generating mechanisms used in the simulation studies, particularly how off-treatment data were generated.

Development of Multiple Imputation Methods Under the Treatment Policy Strategy

The first class of MI methods proposed for handling missing data under a treatment policy strategy is the RB framework. RB MI was the method most frequently identified across the included studies, which may partly reflect its relatively early development compared with other methods.

In 2013, Carpenter et al. [6] introduced the concept of the de facto estimand, defining it as “what would be the effect seen in practice”. This concept is closely aligned with the treatment policy strategy. In that article, an RB MI framework using an MVN model for a longitudinal continuous outcome was introduced, and the implementation of RB MI was discussed under several different RB assumptions, including J2R, LMCF, CIR, and CR. This early discussion likely contributed to the wider awareness and adoption of RB MI in subsequent research. Building on this framework, subsequent studies extended RB MI beyond continuous outcomes. In 2014, Keene et al. [28] extended the method to recurrent event outcomes. Subsequently, Akacha and Ogundimu [26] reformulated it within a continuous-time recurrent event framework, rather than a discrete-time interval count framework with constant event intensity, thus permitting greater flexibility in both imputation and analysis through time-varying baseline intensities. Building on this continuous-time framework, Gao et al. [25] further specified the recurrent event process using a piecewise exponential proportional intensity model with frailty, rather than the constant or Weibull baseline intensity models considered in Akacha and Ogundimu [26]. This allows greater flexibility in capturing time-varying event rates.

In 2017, another adaptation of the RB method for continuous outcomes was introduced by Tang [17]. The article proposed a computationally more convenient algorithm that reformulated imputation for mixed-effects models for repeated measures and pattern mixture models, thereby enabling implementation using existing software such as PROC MI in SAS. Subsequent work extended RB MI from a more conceptual perspective. In particular, White et al. [14] formulated RB MI within a causal modelling framework, while still relying on an MVN model. From a causal perspective, Fang and Jin [16] further developed the implementation of RB MI by proposing a sequential linear regression approach. However, these developments were primarily concerned with continuous outcomes under normality assumptions. Since continuous outcomes may in practice deviate from normality, Liu et al. [13] later extended the RB framework to accommodate non-normal continuous outcomes.

For binary outcomes, extensions of these modelling ideas were identified to proceed in two distinct directions. Some authors retained the modelling framework used for continuous outcomes for imputation using an MVN, but extended the MVN approach to binary data by using adaptive rounding algorithms post-imputation. In contrast, others modelled binary outcomes more directly, either on a latent scale using a multivariate probit model or on the observed binary scale using logistic regression. Specifically, Rehal et al. [21] proposed an MVN model combined with an adaptive rounding algorithm as a technical extension to enable RB MI for binary outcomes. Tang [24] introduced a multivariate probit model under a monotone data augmentation framework, covering CR and J2R assumptions for binary outcomes. Building on this, Lu [20] employed a composite likelihood approach for the multivariate probit model to implement CR and J2R MI. For the logistic regression-based approaches, Gao et al. [23] used logistic regression models as working models to account for the dependence among longitudinal outcomes, with only observations from the control group used under the CR assumption. Tang [24] further applied logistic regression for binary outcomes and additionally a proportional odds model for ordinal outcomes under monotone data augmentation. As discussed in the continuous outcomes section, Fang and Jin [16] also extended this approach to a causal inference perspective, applying a logistic model corresponding to binary outcomes.

However, these existing methods share a common limitation: they were developed exclusively under CR and J2R assumptions, leaving CIR and LMCF, which are equally important RB assumptions, unaddressed for binary outcomes. To fill this gap, Cro et al. [22] formally introduced both the MVN model with adaptive rounding and the multivariate probit model within the RB MI framework, and extended both approaches to accommodate all four RB assumptions, with findings suggesting that the multivariate probit model provides more accurate inference than the MVN model with adaptive rounding.

In 2019, Atkinson et al. [32] extended RB MI to time-to-event outcomes using a Weibull proportional hazards model. Five years later, Jin [29] proposed a separate extension of RB MI under the J2R and CR assumptions for survival analysis with time-dependent covariates using a Cox model.

Beyond the RB framework, both the RTB and washout methods also closely align with the RB concept, where information is borrowed from a “reference” group. The RTB method borrows information from the

baseline mean. Only one RTB method based on the normal distribution for a continuous outcome was identified in our review. Other RTB methods that have been developed have been specified to align with a hypothetical strategy and were therefore excluded according to our eligibility criteria [56, 57]. No extensions of RTB for handling alternative outcome types under a treatment policy strategy were observed within this review.

The washout method was described by Wang et al. [4] as a simplified RB approach in which imputation is performed only at the final visit based on the distribution of control-arm completers. However, when we conducted a separate broad search using the term “washout method”, we found that although this term appeared in some trial reports, it was in fact often used to refer to an RTB-type approach rather than a distinct methodological framework. Moreover, apart from the article by Wang et al. [4], there are only two papers that proposed a placebo-washout method using a non-MI framework.

RD MI was the second class of MI methods proposed for handling missing data under a treatment policy strategy. An important principle in the EMA's 2010 missing data guidance is that stopping study treatment does not imply that outcome collection should also stop [58]. Therefore, outcome data should continue to be collected after treatment discontinuation whenever possible. These post-ICE data can be described as retrieved dropout data. This motivates RD MI, a class of methods that uses observed off-treatment data to impute missing outcomes. Roger et al. [27] anticipated a growing trend towards analyses based on the treatment policy strategy. However, the majority of approaches for handling missing recurrent event outcomes relied on the MAR assumption. This motivated the development of off-treatment imputation approaches for recurrent event outcomes, where missing data following treatment discontinuation are imputed based solely on observed off-treatment data from subjects who discontinued treatment but remained in the trial. Building on the off-treatment imputation idea proposed by Polverejan and Dragalin [59] and Roger et al. [27], Hartley et al. [31] extended the off-treatment imputation framework to accommodate time-to-event endpoints. Around the same period, He et al. [30] extended the retrieved dropout framework to time-to-event endpoints. Unlike Hartley et al. [31], whose work was grounded in the off-treatment imputation concept, He et al. [30] drew upon the retrieved dropout concept originally discussed by McEvoy [60]. Although the two concepts are referred to differently in the literature, they share the same underlying principle of using observed off-treatment data to inform the imputation of missing outcomes. Beyond time-to-event endpoints, RD MI methods have also been extensively developed for continuous outcomes, which represent the most common outcome type in clinical trials. In 2022, Wang and Hu [5] first proposed a last-visit imputation model, where missing outcomes for the final visit only are imputed using separate models for each treatment arm, incorporating the last on-treatment visit, ignoring intermediate outcomes. This was later extended by Wang et al. [8] into a comprehensive implementation utilising all available RD data as a repeated-measure imputation model, which imputes all endpoints, addressing the practical challenge that the original approach required extensive programming derivations. Building on the same principle of incorporating observed off-treatment data into the imputation model under a repeated-measure imputation framework, Drury et al. [10] proposed a sequential MI approach for RD methods using both Historic and Current models. Alternatively, Noci et al. [15] proposed an RD implementation based on an

MVN model, with a different formulation of the Historic and Current models. Collectively, these studies demonstrate that RD MI encompasses a spectrum of implementations, ranging from relatively simple approaches to more complex modelling strategies (e.g., Historic and Current models), depending on how observed off-treatment data are utilised. However, compared with RB MI, the current body of literature on RD MI remains relatively limited, and further extensions to a wider range of outcome types are still needed.

Although both RB- and RD-based approaches provide principled ways to handle missing data under treatment policy estimands, each has its own limitations. RD-based approaches may be impractical when little observed off-treatment data are available, whereas RB-based approaches rely on strong assumptions and do not utilise any observed off-treatment information. To address these limitations, hybrid imputation methods have more recently been proposed to combine the strengths of both approaches. Cro et al. [18] proposed a hybrid method called RD RBI-C. Subsequently, Yuan et al. [12] extended the OICS-R model of RD from Drury et al. [10] to develop the RD_p methods and compared their performance with the RD methods proposed by Drury et al. [10] and Wang and Hu [5]. In addition, rather than using RB MI, Li et al. [19] used RTB and incorporated RD information into the RTB model.

Strengths and Weaknesses of the Study

The strength of this review lies in providing a comprehensive overview of MI methods that can be adopted under the treatment policy strategy. By focusing on methodological papers, this review examines the underlying differences in proposed imputation models, explores how methods vary across different outcome types, and traces the evolution of these methods over time. Furthermore, it discusses key issues within individual classes of imputation methods. Collectively, these findings offer valuable insights and directions to inform future methodological research and support applied researchers in selecting appropriate analysis methods.

However, this review has several limitations. First, from the perspective of search terms, the treatment policy strategy can be described using various terms, but only “de facto” was included as an alternative term. During the review process, “effectiveness estimand” and “ITT estimand” were identified as other potentially relevant terms that were not captured. Second, the inclusion of the term “imputation” in the search strategy may have excluded some early foundational studies, as imputation methods were not always explicitly described as such in titles or abstracts at the time. Consequently, some early published imputation methods that align with the treatment policy strategy may not have been identified. Finally, this review focused exclusively on MI methods developed specifically for the treatment policy strategy. Methods originally developed for other estimand strategies were excluded, even though some may also be applicable under a treatment policy strategy.

Future Research Recommendations

While several different MI methods have been identified that can be used under the treatment policy strategy and comparative studies have begun to assess the performance of the main broad classes of

RB and RD MI approaches, the scope of these evaluations has generally been limited. Recent years have seen the development of several new MI methods under treatment policy strategy, including RD MI method variants based on alternative imputation models and hybrid approaches designed to address the limitations of previously published methods. However, no study has provided a comprehensive head-to-head comparison of established RB, RD, and RTB approaches and recently published hybrid MI methods. Therefore, a systematic comparison across a broader range of MI methods aligning with the treatment policy strategy would be beneficial. Furthermore, although RB methods have been extended to a wide range of outcome types, not all imputation methods under treatment policy have been developed to the same extent. For example, the three novel hybrid methods have only been evaluated for continuous outcomes. This may be partly because most of these methods have been developed relatively recently. Extending these approaches to other outcome types could therefore be a valuable direction for future research.

Conclusion

This review provides a comprehensive overview of MI methods for handling missing data in randomised controlled trials targeting a treatment policy strategy. Through a scoping review of the literature, RB, RD, RTB, washout and hybrid approaches were identified and synthesised across different outcome types. Understanding the assumptions and limitations underlying each method is essential for selecting the most appropriate approach in a given clinical trial setting.

Abbreviations

BOCF: Baseline Observation Carried Forward

CCA: Complete-Case Analysis

CIR: Copy Increments in Reference

CR: Copy Reference

DAR: Deviation at Random

DNAR: Deviation Not at Random

ICEs: Intercurrent Events

J2R: Jump to Reference

LMCF: Last Mean Carried Forward

LOCF: Last Observation Carried Forward

MAR: Missing at Random

MI: Multiple Imputation

MICE: Multiple Imputation by Chained Equations

MMRM: Mixed Model for Repeated Measures

MMRMD: Mixed Model for Repeated Measures Directly at the Composite Score Level

MVN: Multivariate Normal

OICS-R: On/Off-Intercepts Model with Common Slopes Using Residuals

RB: Reference-Based

RD: Retrieved Dropout

RD RBI-C: Retrieved Dropout Reference-Based Centred Multiple Imputation

RTB: Return to Baseline

SCORE-MICE: Multiple Imputation by Chained Equations at the Composite Score Level

TM: Trimmed Mean

UK: United Kingdom

Declarations

Ethics approval and consent to participate

No ethical approval or consent was required for this review of previously published studies.

Consent for publication

Not applicable.

Availability of data and materials

The data extraction form is provided in the supplementary materials. The data extracted and analysed during this review are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' Contributions

MJ conducted the scoping review, designed the data extraction form, extracted and analysed the data, interpreted the results, and drafted and revised the manuscript. SC and LJ supervised the project; provided guidance on uncertainties arising during the literature search, study selection, data extraction, and interpretation of results; contributed to drafting; and critically reviewed and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Infrastructure support for Imperial Clinical Trial Unit was provided by the NIHR Imperial Biomedical Research Centre. The authors would like to thank Jacqueline Kemp, Liaison Librarian (Medicine) at Imperial College London, for her helpful advice on literature search strategy.

Authors' information

Imperial Clinical Trials Unit, Imperial College London, Stadium House, 68 Wood Lane, London, UK

Mingxin Jiang, Leila Janani, Suzie Cro

References

1. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials. 2019.
2. Kahan BC, Hindley J, Edwards M, Cro S, Morris TP. The estimands framework: a primer on the ICH E9 (R1) addendum. *BMJ*. 2024;384. <https://doi.org/10.1136/bmj-2023-076316>.
3. Carpenter JR, Smuk M. Missing data: A statistical framework for practice. *Biometrical J*. 2021;63:915–47. <https://doi.org/10.1002/bimj.202000196>.
4. Wang Y, Tu W, Kim Y, Sinks S, He J, Cambon A, et al. Statistical methods for handling missing data to align with treatment policy strategy. *Pharmaceutical Statistics*. 2023;22:650–70. <https://doi.org/10.1002/pst.2299>.
5. Wang S, Hu H. Impute the missing data using retrieved dropouts. *BMC Med Res Methodol*. 2022;22:82. <https://doi.org/10.1186/s12874-022-01509-9>.
6. Carpenter JR, Roger JH, Kenward MG. Analysis of Longitudinal Trials with Protocol Deviation: A Framework for Relevant, Accessible Assumptions, and Inference via Multiple Imputation. *Journal of Biopharmaceutical Statistics*. 2013;23:1352–71. <https://doi.org/10.1080/10543406.2013.834911>.
7. Zhang Y, Zimmer Z, Xu L, Lam RLH, Huyck S, Golm G. Missing Data Imputation With Baseline Information in Longitudinal Clinical Trials. *Statistics in Biopharmaceutical Research*. 2022;14:242–8. <https://doi.org/10.1080/19466315.2020.1845234>.

8. Wang S, Schwartz PF, Mancuso JP. Comprehensive implementations of multiple imputation using retrieved dropouts for continuous endpoints. *BMC Med Res Methodol*. 2025;25:47. <https://doi.org/10.1186/s12874-025-02494-5>.
9. Bell J, Drury T, Mütze T, Pipper CB, Guizzaro L, Mitroiu M, et al. Estimation Methods for Estimands Using the Treatment Policy Strategy; a Simulation Study Based on the PIONEER 1 Trial. *Pharmaceutical Statistics*. 2025;24:e2472. <https://doi.org/10.1002/pst.2472>.
10. Drury T, Abellan JJ, Best N, White IR. Estimation of Treatment Policy Estimands for Continuous Outcomes Using Off-Treatment Sequential Multiple Imputation. *Pharmaceutical Statistics*. 2024;23:1144–55. <https://doi.org/10.1002/pst.2411>.
11. Martin GP, Jenkins DA, Bull L, Sisk R, Lin L, Hulme W, et al. Toward a framework for the design, implementation, and reporting of methodology scoping reviews. *Journal of Clinical Epidemiology*. 2020;127:191–7. <https://doi.org/10.1016/j.jclinepi.2020.07.014>.
12. Yuan C, Yang Z, Liu J, Li X, Chen B, Han T, et al. An adaptive imputation method of missing data for sparsely retrieved dropouts in treatment policy strategy. *Contemporary Clinical Trials*. 2025;152:107886. <https://doi.org/https://doi.org/10.1016/j.cct.2025.107886>.
13. Liu S, Zhang Y, Golm GT, Liu G (Frank), Yang S. Robust analyzes for longitudinal clinical trials with missing and non-normal continuous outcomes. *Statistical Theory and Related Fields*. 2024;8:1–14. <https://doi.org/10.1080/24754269.2023.2261351>.
14. White I, Joseph R, Best N. A causal modelling framework for reference-based imputation and tipping point analysis in clinical trials with quantitative outcome. *Journal of Biopharmaceutical Statistics*. 2020;30:334–50. <https://doi.org/10.1080/10543406.2019.1684308>.
15. Noci A, Wolbers M, Abt M, Baayen C, Burger HU, Jin M, et al. A Comparison of Estimand and Estimation Strategies for Clinical Trials in Early Parkinson's Disease. *Statistics in Biopharmaceutical Research*. 2023;15:491–501. <https://doi.org/10.1080/19466315.2022.2116476>.
16. Fang Y, Jin M. Sequential modeling for a class of reference-based imputation methods in clinical trials with quantitative or binary outcomes. *Statistics in Medicine*. 2022;41:1525–40. <https://doi.org/10.1002/sim.9303>.
17. Tang Y. An Efficient Multiple Imputation Algorithm for Control-Based and Delta-Adjusted Pattern Mixture Models using SAS. *Statistics in Biopharmaceutical Research*. 2017;9:116–25. <https://doi.org/10.1080/19466315.2016.1225595>.
18. Cro S, Roger JH, Carpenter JR. Handling Partially Observed Trial Data After Treatment Withdrawal: Introducing Retrieved Dropout Reference-Base Centred Multiple Imputation. *Pharmaceutical Statistics*. 2024;pst.2416. <https://doi.org/10.1002/pst.2416>.
19. Li X, Yang Z, Yuan C, Liu J, Li Z. Impute the missing data by combining retrieved dropouts and return to baseline method. *PLOS ONE*. 2025;20:1–16. <https://doi.org/10.1371/journal.pone.0323496>.
20. Lu K. Reference-based pattern-mixture models for analysis of longitudinal binary data. *Stat Methods Med Res*. 2020;29:3770–82. <https://doi.org/10.1177/0962280220941880>.

21. Rehal S, Cro S, Phillips PP, Fielding K, Carpenter JR. Handling intercurrent events and missing data in non-inferiority trials using the estimand framework: A tuberculosis case study. *Clinical Trials*. 2023;20:497–506. <https://doi.org/10.1177/17407745231176773>.
22. Cro S, Quartagno M, White IR, Carpenter JR. Reference-Based Multiple Imputation for Longitudinal Binary Data. *Statistics in Medicine*. 2025;44:e10301. <https://doi.org/10.1002/sim.10301>.
23. Gao F, Liu G, Zeng D, Diao G, Heyse JF, Ibrahim JG. On inference of control-based imputation for analysis of repeated binary outcomes with missing data. *Journal of Biopharmaceutical Statistics*. 2017;27:358–72. <https://doi.org/10.1080/10543406.2017.1289957>.
24. Tang Y. Controlled pattern imputation for sensitivity analysis of longitudinal binary and ordinal outcomes with nonignorable dropout. *Statistics in Medicine*. 2018;37:1467–81. <https://doi.org/10.1002/sim.7583>.
25. Gao F, Liu GF, Zeng D, Xu L, Lin B, Diao G, et al. Control-based imputation for sensitivity analyses in informative censoring for recurrent event data. *Pharmaceutical Statistics*. 2017;16:424–32. <https://doi.org/10.1002/pst.1821>.
26. Akacha M, Ogundimu EO. Sensitivity analyses for partially observed recurrent event data. *Pharmaceutical Statistics*. 2016;15:4–14. <https://doi.org/https://doi.org/10.1002/pst.1720>.
27. Roger JH, Bratton DJ, Mayer B, Abellan JJ, Keene ON. Treatment policy estimands for recurrent event data using data collected after cessation of randomised treatment. *Pharmaceutical Statistics*. 2019;18:85–95. <https://doi.org/10.1002/pst.1910>.
28. Keene ON, Roger JH, Hartley BF, Kenward MG. Missing data sensitivity analysis for recurrent event data using controlled imputation. *Pharmaceutical Statistics*. 2014;13:258–64. <https://doi.org/10.1002/pst.1624>.
29. Jin M. Imputation methods for informative censoring in survival analysis with time dependent covariates. *Contemporary Clinical Trials*. 2024;136:107401. <https://doi.org/10.1016/j.cct.2023.107401>.
30. He J, Crackel R, Koh W, Chen L-W, Li F, Zhang J, et al. Retrieved-Dropout-Based multiple imputation for time-to-event data in cardiovascular outcome trials. *Journal of Biopharmaceutical Statistics*. 2023;33:234–52. <https://doi.org/10.1080/10543406.2022.2118763>.
31. Hartley B, Drury T, Lettis S, Mayer B, Keene ON, Abellan JJ. Estimation of a treatment policy estimand for time to event data using data collected post discontinuation of randomised treatment. *Pharmaceutical Statistics*. 2022;21:612–24. <https://doi.org/10.1002/pst.2189>.
32. Atkinson A, Kenward MG, Clayton T, Carpenter JR. Reference-based sensitivity analysis for time-to-event data. *Pharmaceutical Statistics*. 2019;18:645–58. <https://doi.org/10.1002/pst.1954>.
33. Cro S, Morris TP, Roger JH, Carpenter JR. Comments on ‘standard and reference-based conditional mean imputation’: Regulators and trial statisticians be aware! *Pharmaceutical Statistics*. 2024;23:598–603. <https://doi.org/10.1002/pst.2373>.
34. Wolbers M, Noci A, Delmar P, Gower-Page C, Yiu S, Bartlett JW. Standard and reference-based conditional mean imputation. *Pharmaceutical Statistics*. 2022;21:1246–57.

- <https://doi.org/10.1002/pst.2234>.
35. Cro S, Carpenter JR, Kenward MG. Information-Anchored Sensitivity Analysis: Theory and Application. *Journal of the Royal Statistical Society Series A: Statistics in Society*. 2019;182:623–45. <https://doi.org/10.1111/rssa.12423>.
 36. Yang S, Zhang Y, Liu GF, Guan Q. SMIM: A unified framework of survival sensitivity analysis using multiple imputation and martingale. *Biometrics*. 2023;79:230–40. <https://doi.org/10.1111/biom.13555>.
 37. Quan H, Qi L, Luo X, Darchy L. Considerations of multiple imputation approaches for handling missing data in clinical trials. *Contemporary Clinical Trials*. 2018;70:62–71. <https://doi.org/10.1016/j.cct.2018.05.008>.
 38. Bartlett JW. Reference-Based Multiple Imputation—What is the Right Variance and How to Estimate It. *Statistics in Biopharmaceutical Research*. 2023;15:178–86. <https://doi.org/10.1080/19466315.2021.1983455>.
 39. Cai X, Gewandter JS, He H, Turk DC, Dworkin RH, McDermott MP. Estimands and missing data in clinical trials of chronic pain treatments: advances in design and analysis. *Pain*. 2020;161:2308–20. <https://doi.org/10.1097/j.pain.0000000000001937>.
 40. Yan M, Zhou L, Zhao C. Comparison of Different Approaches in Handling Missing Data in Longitudinal Multiple-Item Patient-Reported Outcomes: A Simulation Study. *Health and Quality of Life Outcomes*. 2025;23:34. <https://doi.org/10.1186/s12955-025-02364-0>.
 41. Medcalf E, Turner RM, Espinoza D, Zhu L, Bell KJL. Evaluating missing outcome data methods for the analysis of the MEL-SELF trial of patient-led surveillance for early-stage melanoma: A simulation study. *Contemporary Clinical Trials*. 2026;161:108186. <https://doi.org/https://doi.org/10.1016/j.cct.2025.108186>.
 42. Lu K, Li D, Koch GG. Comparison Between Two Controlled Multiple Imputation Methods for Sensitivity Analyses of Time-to-Event Data With Possibly Informative Censoring. *Statistics in Biopharmaceutical Research*. 2015;7:199–213. <https://doi.org/10.1080/19466315.2015.1053572>.
 43. Floden L, DeRosa M, Roydhouse J, Beaumont JL, Hudgens S. A demonstration of estimands and sensitivity analyses for time-to-deterioration of patient reported outcomes. *Journal of Biopharmaceutical Statistics*. 2025;35:918–32. <https://doi.org/10.1080/10543406.2024.2341649>.
 44. Lu K. An analytic method for the placebo-based pattern-mixture model. *Statistics in Medicine*. 2014;33:1134–45. <https://doi.org/https://doi-org.iclibezp1.cc.ic.ac.uk/10.1002/sim.6008>.
 45. Tang Y. Short notes on maximum likelihood inference for control-based pattern-mixture models. *Pharmaceutical Statistics*. 2015;14:395–9. <https://doi.org/https://doi-org.iclibezp1.cc.ic.ac.uk/10.1002/pst.1698>.
 46. Mehrotra DV, Liu F, Permutt T. Missing data in clinical trials: control-based mean imputation and sensitivity analysis. *Pharmaceutical Statistics*. 2017;16:378–92. <https://doi.org/https://doi-org.iclibezp1.cc.ic.ac.uk/10.1002/pst.1817>.

47. García-Hernandez A, Pérez T, Pardo MDC, Rizopoulos D. A Flexible Analytical Framework for Reference-Based Imputation, Delta Adjustment and Tipping-Point Stress-Testing. *Statistics in Biopharmaceutical Research*. 2023;15:781–93. <https://doi.org/10.1080/19466315.2022.2151506>.
48. Lou J, Rettiganti M, Qu Y. Direct Estimation for Commonly Used Pattern-Mixture Models in Clinical Trials. *Statistics in Biopharmaceutical Research*. 2024;:1–13. <https://doi.org/10.1080/19466315.2024.2416417>.
49. Liu S, Yang S, Zhang Y, Liu G (Frank). Sensitivity analyses in longitudinal clinical trials via distributional imputation. *Stat Methods Med Res*. 2023;32:181–94. <https://doi.org/10.1177/09622802221135251>.
50. García-Hernandez A, Pérez T, del Carmen Pardo M, Rizopoulos D. An illness–death multistate model to implement delta adjustment and reference-based imputation with time-to-event endpoints. *Pharmaceutical Statistics*. 2024;23:219–41. <https://doi.org/https://doi-org.iclibezp1.cc.ic.ac.uk/10.1002/pst.2348>.
51. Fairfax SR, Yang S. Distributional imputation for the analysis of censored recurrent events. *Statistics in Medicine*. 2024;43:2622–40. <https://doi.org/https://doi-org.iclibezp1.cc.ic.ac.uk/10.1002/sim.10087>.
52. Kenward MG. Controlled multiple imputation methods for sensitivity analyses in longitudinal clinical trials with dropout and protocol deviation. *Clinical Investigation*. 2015;5:311–20. <https://doi.org/10.4155/cli.14.132>.
53. Cro S, Morris TP, Kenward MG, Carpenter JR. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide. *Statistics in Medicine*. 2020;39:2815–42. <https://doi.org/10.1002/sim.8569>.
54. Medcalf E, Turner RM, Espinoza D, He V, Bell KJL. Addressing missing outcome data in randomised controlled trials: A methodological scoping review. *Contemporary Clinical Trials*. 2024;143:107602. <https://doi.org/10.1016/j.cct.2024.107602>.
55. Liu GF, Pang L. Control-Based Imputation and Delta-Adjustment Stress Test for Missing Data Analysis in Longitudinal Clinical Trials. *Statistics in Biopharmaceutical Research*. 2017;9:186–94. <https://doi.org/10.1080/19466315.2016.1256830>.
56. Jin M. A hybrid return to baseline imputation method to incorporate MAR and MNAR dropout missingness. *Contemporary Clinical Trials*. 2022;120:106859. <https://doi.org/https://doi.org/10.1016/j.cct.2022.106859>.
57. Qu Y, Dai B. Return-to-baseline multiple imputation for missing values in clinical trials. *Pharmaceutical Statistics*. 2022;21:641–53. <https://doi.org/https://doi-org.iclibezp1.cc.ic.ac.uk/10.1002/pst.2191>.
58. European Medicines Agency. Guideline on Missing Data in Confirmatory Clinical Trials. London, UK: European Medicines Agency; 2010.
59. Polverejan E, Dragalin V. Aligning Treatment Policy Estimands and Estimators—A Simulation Study in Alzheimer’s Disease. *Statistics in Biopharmaceutical Research*. 2020;12:142–54.

60. McEvoy BW. Missing data in clinical trials for weight management. Journal of Biopharmaceutical Statistics. 2016;26:30–6. <https://doi.org/10.1080/10543406.2015.1094814>.

Figures

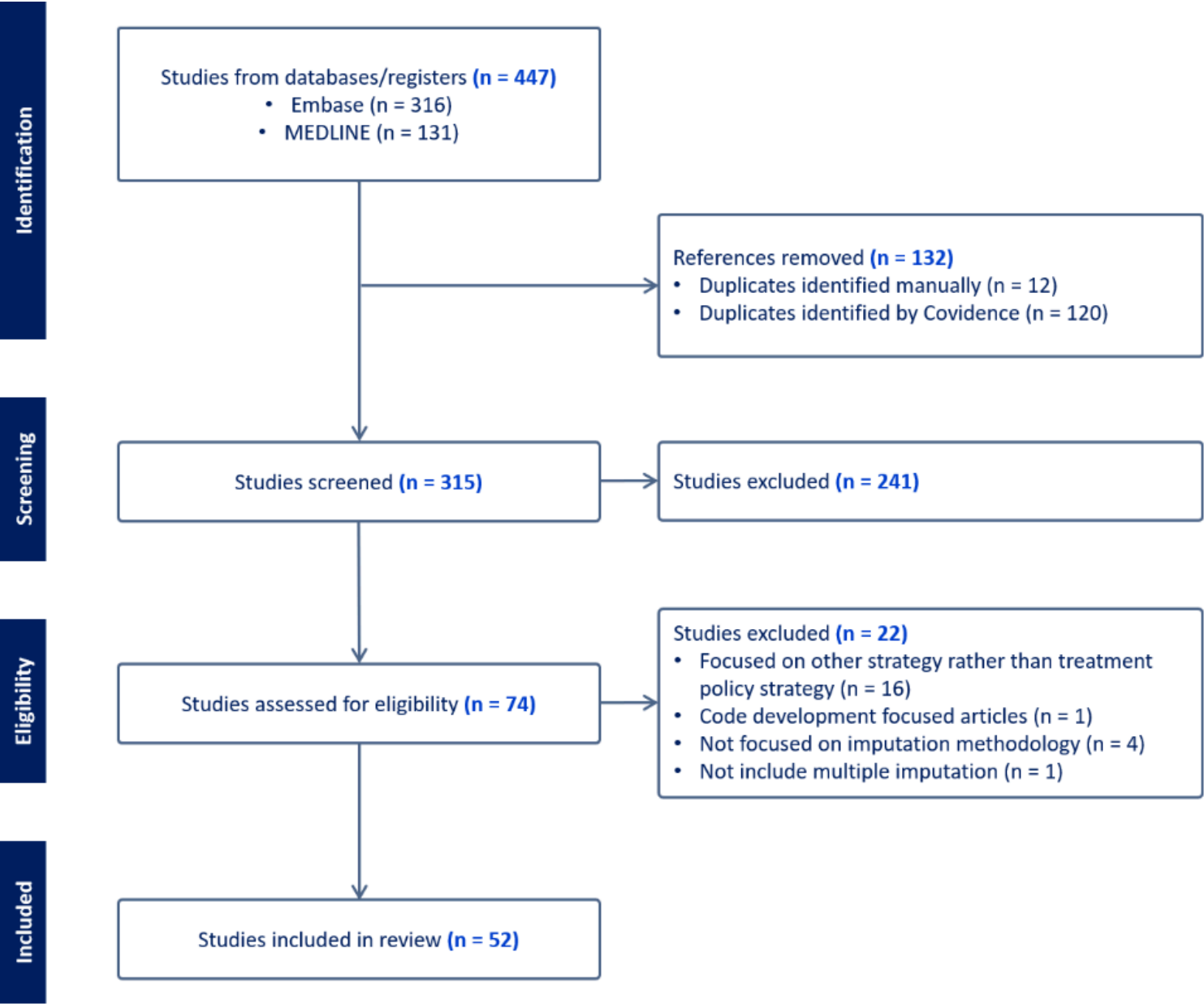


Figure 1

PRISMA flow diagram

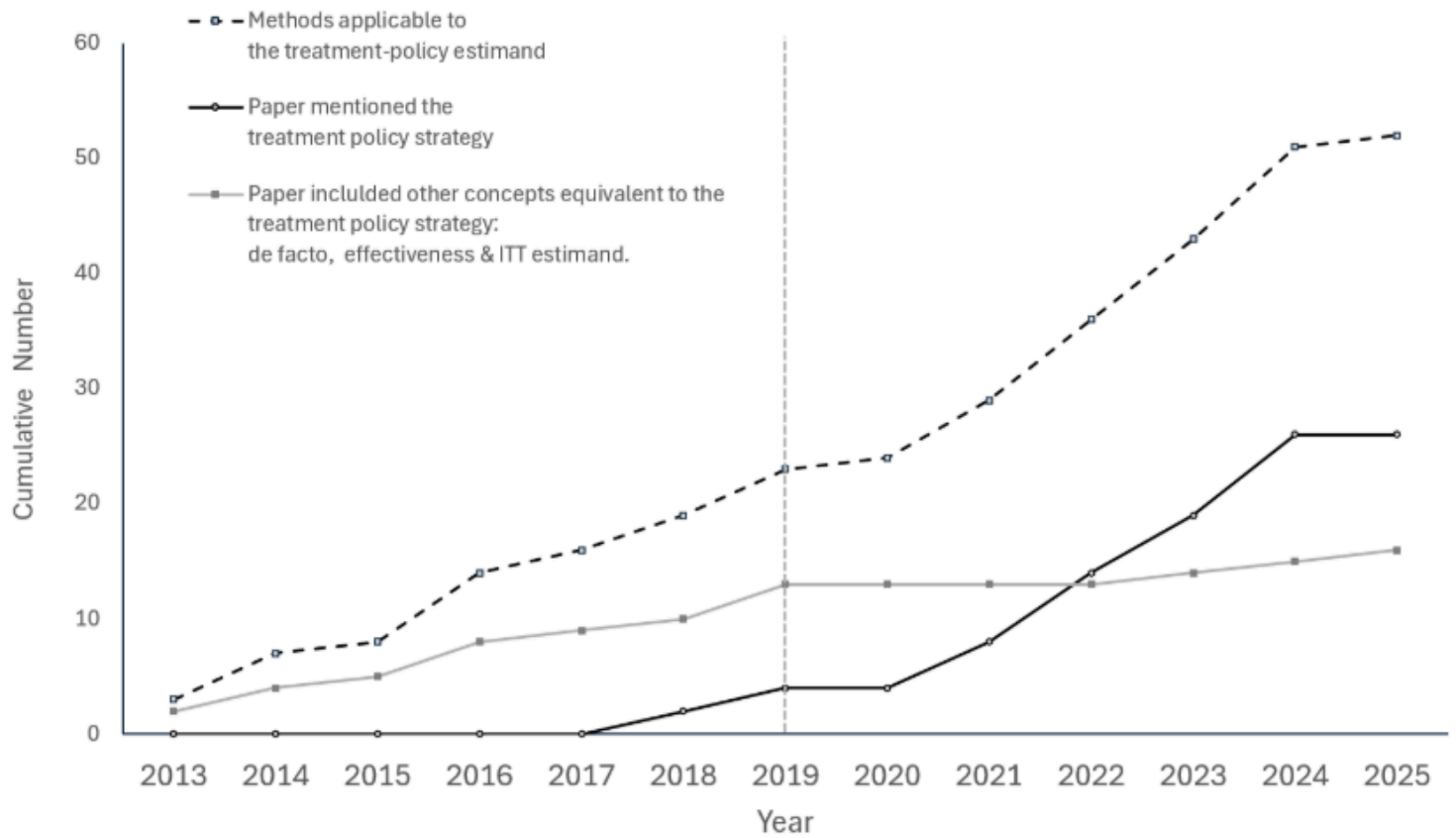


Figure 2

Cumulative numbers of publications by year for methods applicable to the treatment-policy estimand. The dashed vertical line indicates the publication year of the ICH E9 (R1) addendum.

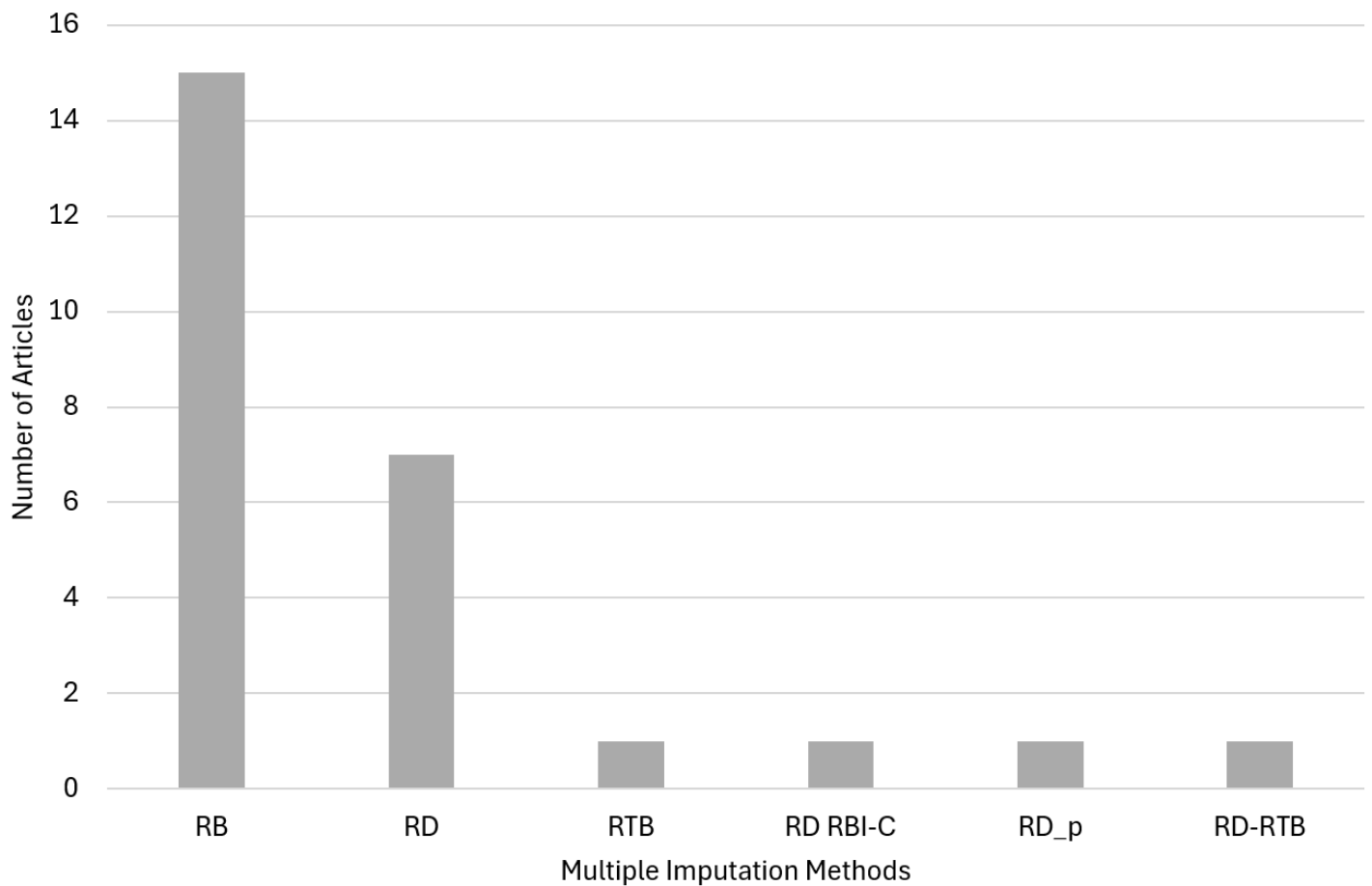


Figure 3

Summary of imputation methods across articles that provided new multiple imputation methods

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1PRISMAchecklistforreportingsystematicreviews.docx](#)
- [Additionalfile2MoreinformationonStudyIdentificationandSelection.docx](#)
- [Additionalfile3DataExtractionSheet.docx](#)