Dealing with Missing Data in Real-World Data: A Scoping Review of Simulation Studies

Alejandra E Guevara Morel (a.e.guevaramorel@vu.nl)
Vrije Universiteit Amsterdam, Amsterdam Public Health Research Institute

Anita N Varga
Vrije Universiteit Amsterdam, Amsterdam Public Health Research Institute

Martijn W Heymans
Amsterdam UMC Location VUmc

Johanna M Dongen
Vrije Universiteit Amsterdam, Amsterdam Public Health Research Institute

Digna JF Schaik
Amsterdam UMC Location VUmc

Maurits W Tulder
Vrije Universiteit Amsterdam

Judith E Bosmans
Vrije Universiteit Amsterdam, Amsterdam Public Health Research Institute

Research Article

Keywords: missing data, real-world data, observational data, simulation

Posted Date: May 19th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1619388/v1

RESEARCH

Dealing with Missing Data in Real-World Data: A Scoping Review of Simulation Studies

Alejandra E Guevara Morel[1*], Anita N Varga[1], Martijn W Heymans[2], Johanna M van Dongen[1,3], Digna JF van Schaik[4,5], Maurits W van Tulder[6] and Judith E Bosmans[1]

[*]Correspondence: a.e.guevaramorel@vu.nl
[1] Department of Health Sciences, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam Public Health Research Institute, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
[2]Department of Epidemiology and Data Science, Amsterdam University Medical Center, Amsterdam, The Netherlands
[3]Department of Health Sciences, Faculty of Science, Vrije Universiteit Amsterdam, MOVE Research Institute, Amsterdam, The Netherlands
[4]Amsterdam UMC location Vrije Universiteit Amsterdam, Department of Psychiatry, Amsterdam, The Netherlands
[5]Amsterdam Public Health, Mental Health program, Amsterdam, The Netherlands
[6]Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam Movement Science Research Institute, Amsterdam, The Netherlands

Abstract

BACKGROUND: Real-world data are increasingly being used for as a complement to randomized controlled trials (RCTs) for evaluating the effectiveness and cost-effectiveness of healthcare interventions. Real-world data often are expected to have higher generalizability by including more representative patient populations and resembling daily clinical practice better than RCTs. However, since data are not collected for research purposes, missing observations in real-world data are highly common. Inadequate handling of missing observations may lead to biased estimates and invalid conclusions. The aim of this scoping review was to identify and critically appraise statistical methods for dealing with missing observations in real-world data.

METHODS: We searched PubMed for simulation studies that assessed the performance of statistical methods for dealing with missing observations in real-world data published between January 2000 and December 2018. We searched for simulation studies because well-developed simulation studies may be preferable to choose the best missing data method for real-world data. Information was extracted on the aims of the studies, data generating mechanisms, assessed statistical methods for dealing with missing observations, performance of the statistical method, statistical software used, and authors' remarks and conclusions on the validity and usability of the statistical methods.

RESULTS: Fifty-two studies were eligible for inclusion; 19 assessed methods for missing covariates, 13 for missing outcome(s), 15 for missing observations in both covariates and outcome(s), and in 5 studies it was unclear whether missing observations were present in covariates and/or outcome(s). Eleven studies took into account the multilevel structure of the data, whereas others did not. When imputing single-level missing at random (MAR) data, MICE and multivariate normal imputations (MVNI) seemed the best performing methods. When dealing with multilevel MAR data, multiple imputation-based methods seem to be the most flexible and best performing methods. For data missing not at random (MNAR, 16 studies), selection models and pattern mixture models appear to be most promising.

CONCLUSIONS: The choice of a statistical method to deal with missing observations depends on the type of missing variables and the assumed missing data mechanism. Although MAR is the most commonly assumed missing data mechanism, data that are missing not at random (MNAR) is also common in real-world data, although only few studies evaluated methods for MNAR data.

Keywords: missing data; real-world data; observational data; simulation

1 Background

Historically, randomized controlled trials (RCTs) are considered the "gold standard" for evaluating the effectiveness and cost-effectiveness of healthcare interventions, because their internal validity is high due to the randomization of participants into two or more study groups [1, 2, 3, 4, 5, 6]. However, the external validity of RCTs can be compromised due to strict inclusion criteria resulting in homogeneous study populations in combination with strict treatment protocols [1, 2, 3, 4, 5, 6]. A possible addition to RCTs is the use of data collected in a routine care environment, also known as real-world data [6, 7].The advantage of using real-world data is that they are collected under less strictly controlled circumstances, leading to increased generalizability compared with RCTs [3]. Because of this reason, over the last years real-world data are increasingly being used as a complement to RCTs and play an increasingly important role in generating timely evidence on the (cost-) effectiveness of healthcare interventions in actual clinical practice [2, 4, 6, 7, 8, 9, 10, 11]. Moreover, real world data may be especially suitable for cost-effectiveness analyses, because these studies aim to directly influence reimbursement decisions and therefore should reflect actual clinical practice as much as possible.

Despite the advantages of real-world data studies, there are important methodological pitfalls compared to RCTs [1, 4, 8]. Probably one of the most important pitfalls is that of missing data [4, 12]. Missing observations are a particular concern in real-world data studies, because real-world data are typically collected to support the healthcare process and not for research purposes [8]. Additionally, the complexity of real-world data (e.g., structured data such as diagnosis based on codes in combination with textual or narrative data such as attending physician's notes), the number or rate (i.e., percentage) of missing observations (e.g., which can be higher than in RCTs where reminders are sent or interviews are taken), as well as the reasons for the data to be missing (e.g., missings due to lack of collection or lack of documentation) makes dealing with missing observations in real-world data challenging [11, 13, 14].

Research indicates that missing observations are often inadequately handled in clinical studies potentially leading to biased results and decreased precision, which may in turn result in invalid conclusions [15, 16, 17, 18, 19, 20]. This problem may be even more pronounced when using real world data, where there is a risk of additional bias if missing observations occur in variables that are used to define the cohort or if linkage of all participants is not possible due to missing observations [14]. Therefore, the selection of appropriate methods to handle missing observations in real-world data studies is especially vital.

Traditionally, missing observations have been handled by simply deleting cases with missing values, also known as a complete-case analysis [18, 20]. Other so-called "simple" methods for dealing with missing observations include mean imputation, hot-deck imputation, and last observation carried forward [21, 22]. All of these methods, however, may result in biased estimates and underestimate the variance surrounding these estimates [21]. During the previous decades, more advanced

methods such as maximum likelihood-based methods and multiple imputation, have been developed to overcome the disadvantages of these "simple" methods [20, 21].

Several reviews on how to handle missing observations in epidemiological studies have been previously published. However, the majority of these reviews focused on RCT data [15, 16, 17, 18, 19]. Considering the increased use of real-world data in healthcare research and the additional challenges associated with the presence of missing observations in such data, the aim of this scoping review was to identify and critically appraise available statistical methods for dealing with missing observations in real-world data.

2 Methods

In this scoping review, we systematically searched PubMed for simulation studies that assessed the performance of statistical methods for dealing with missing observations in real-world data. We decided to specifically focus on simulation studies because well developed simulation studies may be preferable to choose the best missing data method for eal-world data. We also decided to restrict our search to PubMed, because we aimed to identify methods for dealing with missing observations in biomedical research.

2.1 Search Strategy

We searched for relevant studies using a combination of terms related to the type of data (e.g., "real-world data", "routinely collected data", "electronic health records") and the methodological aspect to be addressed (e.g., "missing data", "missing observations", "censored values") in the title, abstract, MeSH headings or keywords. The search was restricted to studies published between January 2000 and December 2018 to prevent identification of studies evaluating obsolete methods (more advanced methods like multiple imputation by chained equations only came into use after the year 2000). The full PubMed search strategy is available in Appendix 1. The PubMed search was complemented by searching reference lists of included studies. During the search, a logbook was kept with the date of the search, the keywords used and the results of each search. Titles and abstracts of the retrieved studies were stored in an online electronic library using EndNote X8.

2.2 Study Selection

Two researchers (AEGM and ANV) independently screened titles and abstracts identified by the search for possible inclusion in the study. Studies solely evaluating the performance of "simple" methods, such as deletion methods (e.g., complete-case analysis, available-case analysis) and single imputation techniques (e.g., mean imputation, single regression imputation, hot-deck imputation, last observation carried forward) were excluded [21, 22]. This was done, because previous reviews have already shown that these methods can result in severely biased estimates [18]. Studies were included if they conducted a simulation study to assess the relative performance of two or more statistical methods for dealing with missing observations in real-world data. The focus of the review was the comparison of different methods, therefore studies that exclusively compared variations of the same method were

excluded. In addition, studies had to be published as full papers and had to be written in English. Because the review is focused on real-world data, we excluded studies focusing on specific types of data (e.g., high dimensional data such as genetic and imaging data) and/or experimental study designs (e.g., randomized controlled trials). Additionally, we focused specifically on handling missing data when using real-world data patient-level. Therefore, we excluded methods for handling missing observations in meta-analyses. Finally, survival-based methods were also excluded because our interest did not lie in time-to-event outcome variables, but in estimating mean differences. In case of disagreements about the eligibility of the identified studies between the two reviewers, consensus was sought during a face-to-face meeting. If consensus could not be reached, a third reviewer (JEB) was consulted. Full texts of studies that definitely or possibly fulfilled the inclusion criteria were retrieved. A final decision on the inclusion of a study was based on the full text using the same strategy as described above.

Missing Data Mechanisms

All statistical methods for dealing with missing observations rely on assumptions regarding the relationship between the observed and unobserved (i.e., missing) variables, also known as the *missing data mechanism* [23].Three missing data mechanisms are distinguished; missing at random (MAR), missing completely at random (MCAR) and missing not at random (MNAR).

Missing at random (MAR): this mechanism is assumed by most statistical methods and implies that the probability of "missingness" (i.e., the probability of an observable data point being missed) depends (i.e., is conditional) on observed variables and can be explained by the information available in these variables [16, 18, 21, 24, 25, 26, 27]. However, it is important to bear in mind, that the MAR assumption is fundamentally untestable, because part of the data is not observed [28].

Missing completely at random (MCAR): assumes that the missingness of data is not related to any observed or unobserved variables, thus the probability of missingness is equal for all variables [21]. If the MCAR assumption is met, results may be unbiased. However, this is rarely the case, making MAR a more plausible and less restrictive assumption for missing observations [21, 24, 25, 26, 27, 29].

Missing not at random (MNAR): the probability of missing observations is assumed to depend on unobserved data, making this a highly challenging and problematic mechanism to appropriately deal with because the reasons of missingness are unknown [21].

Box 1: Missing data mechanisms

2.3 Data Extraction

One reviewer (AEGM) independently extracted data from the included studies using a standardized data extraction form, which was based on the structured approach for designing and reporting simulation studies by Morris et al. [30]. Data were extracted on the: [1] aim of the study, [2] data generating mechanisms: characteristics of the data (i.e., simulated data only or a combination of simulated data validated with an applied study), sample size, number of simulated data sets, types of missing variables (i.e., missing observations in covariates and/or outcome variables), percentage of missing observations, assumed missing data mechanism (i.e., missing completely at random (MCAR), missing at random (MAR) and/or missing not at random (MNAR), see Box 1) [24], presence of hierarchical levels in the data (i.e., single-level or multilevel, see Appendix 2), [3] assessed statistical methods for dealing with missing observations, [4] reported performance measures (e.g., bias, empirical standard error, model-based standard error, (root) mean square error, coverage, convergence, computational speed), [5] authors' remarks on the compared methods (e.g., advantages and disadvantages), [6] statistical software used, [7] authors' conclusions on the validity of and recommendations about the usability of the method. Five percent of the extracted studies were randomly selected for a second reviewer (ANV or JMvD) to independently assess the data extraction. Inconsistencies found between the two reviewers were settled in a face-to-face meeting.

2.4 Data Synthesis

The results of the scoping review were narratively described. The methods used for dealing with missing observations were described separately for studies that evaluated single-level and multilevel methods. Methods were further categorized according the used mechanism to account for missing observations. An overview of the identified studies per category and per method was included. General information (e.g., description) on methods that were described in a small number of included studies, was complemented with other sources (e.g.,textbooks). Relevant advantages and disadvantages regarding performance of the methods based on remarks by the authors of the included studies were described.

3 Results

We identified 6,820 potentially eligible studies via our electronic search. After removing 109 duplicates, 6,711 studies were screened on title and abstract, resulting in the exclusion of 6,417 papers. After full-text screening of the remaining 294 studies, 48 studies were included in the scoping review. In addition, four studies [31, 32, 33, 34] were added from our own database (Figure 1). The main reasons for exclusion during full-text screening were that studies had a different aim than the evaluation of the relative performance of methods for dealing with missing observations (n=100) and that studies did not conduct a simulation study to assess the performance of the included methods (n=36, see Figure 1 for other exclusion reasons). Supplementary Table 1 gives an overview of the characteristics of the included studies.

Of the included studies, 42 studies (81%) dealt with missing observations in single-level data and 11 studies (21%) with missing observations in multilevel data. The

majority of studies compared methods assumed a MAR mechanism (47 studies, 90%), and 32 studies (62%) tested the performance of the methods under the assumption of more than one missing data mechanism (i.e., MCAR, MAR and/or MNAR). Nineteen studies (37%) assessed methods for dealing with missing covariates, 13 studies (25%) for missing outcome(s), and 15 studies (29%) for missing observations in both covariates and outcome(s). In 5 studies (10%), it was unclear whether missing observations were present in covariates and/or outcome(s). In 49 studies (94%), the methods' performance was assessed in terms of empirical o model-based standard errors, in 37 studies (71%) in terms of convergence, in 33 studies (63%) in terms of bias, in 25 studies (48%) in terms of coverage, and in 19 studies (37%) in terms of (root) mean squared error(RMSE). Supplementary Table 2 gives an aggregated summary of the simulated data generating mechanisms of the included studies.

3.1 Identified Statistical Methods for Dealing with Missing Data Single-level Data

In total, 42 studies were identified that evaluated five different categories of methods to account for missing observations in single-level data (see Table 1 for the methods). These categories are: 1) deletion methods (27/42 studies, 64%), 2) single imputation methods (2/42 studies, 5%), 3) multiple imputation methods (41/42 studies, 98%), 4) maximum likelihood-based methods (6/42 studies, 14%), and 5) weighting methods (7/42 studies, 17%).

Deletion Methods

Two different deletion methods were identified for dealing with missing observations in single-level data: 1) complete case analysis (CCA) and 2) available case analysis (ACA). CCA (25/42 studies) and ACA (3/42 studies) were mainly used as a reference method against which more advanced imputation methods were compared. The use of complete cases only results in a decreased sample size and, thus, a reduction of statistical power. Consequently, standard errors of the estimates are underestimated in comparison with a full data analysis [35]. Although CCA and ACA will not result in bias under a MCAR mechanism, deletion methods will result in biased estimates if the MCAR assumption does not hold [21, 36, 37, 38]. There is one exception to the latter, that is when only outcome data are missing and it is possible to adjust at baseline for the same outcome variable [39]. An advantage of both CCA and ACA is that many standard statistical methods, such as ordinary least squares (OLS) regression, can be easily applied to complete/available cases only [36]. An advantage of CCA over ACA is that the same dataset is used for all analyses, whereas with ACA the sample may differ between outcome measures [21, 36]. ACA, on the other hand, has a better performance (i.e., less biased effect estimates) than CCA in situations where the data have a multivariate normal distribution [21].

Single Imputation Methods

Two different single imputation methods were assessed for dealing with missing observations in single-level data: 1) mean imputation (2/42 studies), and 2) regression imputation (1/42 studies). Single imputation methods may lead to overly

precise results in two ways. First, the uncertainty about the true value of the missing observation is not taken into account, and second, the variability of the missing observations is not accounted for. As a result, the total uncertainty surrounding the estimates is underestimated, which in turn leads to an increased chance of a Type II error [21, 36, 40, 41, 42]. **Mean imputation** distorts the statistical variability of imputed values and the correlation structure between variables due to the insertion of values that are uncorrelated to the observed values [21, 36, 40, 43]. On the other hand, **regression imputation**, which can also be considered weighted mean imputation, has the disadvantage of artificially strengthening the correlations between variables. This is because the imputed values are predicted based on the correlations between the variable with missing observations and the outcome variable [21, 22, 36, 40, 41, 42]. When the MCAR assumption is violated, both methods result in biased estimates [21, 36].

Multiple Imputation Methods

Two different single-level multiple imputation methods were identified: 1) multivariate normal imputation (MVNI), also referred to as joint modelling (JM) (8/42 studies), and 2) multiple imputation by chained equations (MICE), also referred to as fully conditional specification (FCS) (16/42 studies).

The main advantage of multiple imputation methods is that they take imputation uncertainty into account; that is, multiple imputation methods allow for uncertainty in the estimated values [21, 44]. Standard multiple imputation methods such as MVNI and MICE, rely on a MAR assumption [45, 46]. If data are indeed MAR, then multiple imputation results in unbiased estimates [45]. Multiple imputation gives more valid standard errors than single imputation methods, because uncertainty within and between imputations is accounted for [45]. When handling missing observations in longitudinal data, both MVNI and MICE can be used by including the repeated measurements as separate variables in the imputation model, also referred to as "just another variable" approach (JAV)) [46, 47]. However, this approach does not take into account temporal trends (i.e., longitudinal changes over time) [46, 48]. MVNI and MICE were found to perform similarly well when there are relatively few time points in the data (e.g., 3 time points) and a small number of variables [46]. However, when there are large numbers of variables as well as time points, both methods may experience non-convergence due to over-fitting of the imputation model and/or collinearity between predictor variables [46, 47, 48]. Finally, both MVNI and MICE perform poorly when there are interaction terms in the analysis model that need to be accounted for in the models [44].

Multivariate normal imputation (MVNI) or **joint modelling (JM)**, imputes missing observations by specifying a joint imputation model that assumes a multivariate normal distribution of all variables with missing observations [44, 45, 46, 47, 48]. MVNI has two main limitations that are both related to the fact that MVNI assumes specific parametric models that might not be a perfect fit to the complex data structures found in real-world data [45]. First, MVNI cannot account for non-normally distributed data, which can lead to biased estimates

[45, 48, 49, 50] and poor coverage rates, if data are not normally distributed [44]. To deal with non-normal data when using MVNI, the included studies suggested two MVNI log-transformations of non-normal variables: 1) MVNI with a simple log transformation; and 2) MVNI with a zero-skewness log transformation [44, 45]. Lee et al. concluded that in comparable simulation scenarios, MVNI when compared to MICE both with log-transformations of non-normal variables results in similar performance [45]. However, use of log-transformations for MVNI as a comparison to MICE was assessed in only two studies [44, 45]. A second limitation of MVNI is that due to the multivariate normality assumption, dealing with binary and categorical variables is difficult [44]. When dealing with binary variables, two variations of MVNI were suggested: 1) multiple imputation using MCMC-multivariate normal-coin flipping-nonmonotone pattern (MCMC-D); and 2) multiple imputation using MCMC-multivariate normal-adaptive rounding-nonmonotone pattern (MCMC-A) [51]. When dealing with categorical data, MVNI with rounding imputation to the nearest integer value was suggested [50]. However, the included studies showed that all three methods resulted in biased estimates [50, 51].

Multiple imputation by chained equations (MICE), also known as **fully conditional specification (FCS)**, is a semi-parametric Markov Chain Monte Carlo (MCMC) method. MICE is characterized by its flexibility in allowing for the specification of different regression models for different types of variables [44, 52, 53]. Consequently, MICE can be used with missing observations in continuous and both ordinal and nominal categorical data [40, 44, 48]. Finally, although a MAR mechanism is generally assumed when using MICE, MICE can also be used when data is MNAR [54]. The main difference between MICE and MVNI is that MICE does not rely on a multivariate normality assumption [21, 44, 45, 50, 53, 55].

In order to increase the plausibility of a MAR assumption to hold, MICE allows for the inclusion of additional variables in the imputation model, also known as auxiliary variables [52]. These auxiliary variables do not necessarily need to be part of the analysis model, but should have a strong association with the variables with missing observations and with the missingness of these variables [52]. The addition of such auxiliary variables is possible, because the imputation and analysis models are separated from each other when using MICE [56]. By including auxiliary variables that explain the missingness of data, the "prediction" of missing values is improved, because more data are available to inform the imputations. It is important to bear in mind that because full conditional distributions (i.e., regression models) are specified directly, the outcome has to be explicitly included in the imputation models [56]. An important prerequisite of MICE is that the imputation model should at the minimum include all variables and relationships (e.g., interactions) that are included in the analysis model to ensure consistent estimates [21, 44, 52, 55]. If there is incompatibility between imputation and analysis models, MICE may result in biased estimates [52, 57].

MICE has two main limitations. First, the traditional implementation of MICE fails to account for non-normal (i.e., skewed) distributions of variables with missing

observations, which may lead to substantial bias [44, 47]. In total, four variations of MICE were suggested to impute non-normally distributed data: 1) MICE with predictive mean matching (MICE-PMM); 2) MICE based on classification and regression trees (MICE-CART); 3) log transforming the data [44]; and 4) MICE with polytomous regression (POLY) [58]. The first approach, MICE-PMM, relaxes parametric assumptions and bases the imputation of the missing observations on the observed data distributions [21, 44, 47, 49, 59, 60, 61, 62]. MICE-PMM obtains the assumed distribution for the missing observations by using a "pool" of donors from the observed data [49, 60]. These donors are observations from cases that have similar characteristics/values (i.e., closest predicted values) as cases with a missing observation. Subsequently, a random value from these donors for the missing observation is drawn to impute the missing observation [21]. Therefore, MICE-PMM guarantees the preservation of the range of values in the observed data, the distribution of the data and the associations in the data, and, thus, the plausibility of the imputed values [21, 49, 60]. Additionally, MICE-PMM has the advantage that it is robust to misspecification of the distribution underlying the imputation model [21].

A second non-parametric approach, MICE-CART, is flexible enough to fit non-normal distributions because it uses classification and regression trees as conditional (i.e., regression) models for the imputation [58, 61, 63]. Classification trees are used for categorical variables, and regression trees are used for continuous variables [21, 63]. In a nutshell, classification and regression trees (CART) models look for predictors in the data and then create cut-off points in these predictors, which are used to divide the sample (i.e., the available data) [21]. The different samples that result from these splits are expected to be more homogeneous [21]. The splitting process is repeated in each of the sub-samples until the tip of the "branch" only has two options. It is important to take into consideration that when there are too many answer categories (i.e., levels) in a categorical variable, MICE-CART may become computationally intensive [63]. A potential disadvantage of MICE-CART is that it is prone to over-fitting the data because it uses the observed data only [63].

A third approach is to use log-transformation, resulting in a normal distribution of the data after which the missing observations can be imputed [21, 44]. After imputing the data, values need to be back-transformed to the original scale in order to perform the analysis [44]. However, log transforming data often only partially resolves non-normality [44]. Additionally, log-transforming data might not achieve normality of the data, on the contrary it can lead to anomalies in the distribution of the data (e.g., by altering the mean and variability of the data) and/or can distort the relationships between variables [21, 44]. Therefore, this method is generally not recommended. A fourth approach is to use MICE with polytomous regression (POLY). However, this method was assessed only in one study where it was shown to have a higher performance when imputing nominal variables compared to basic MICE and MICE-CART [58].

A second limitation of MICE is that it does not take into account temporal trends in

longitudinal missing time-varying covariates, specifically when the number of variables as well as time points is large (e.g., 5 or more time points and/or 3 or more variables with missing observations) [46, 48]. One variation of MICE was suggested that deals with time-varying covariates: two-fold FCS [46, 47, 48, 53, 64]. Two-fold FCS takes the longitudinal data structure into account by imputing values in a variable at a specific time point, using only data from that specific time point and the immediately adjacent ones [46, 47]. This prevents over-fitting and collinearity issues and hence prevents convergence issues [46, 53, 64]. By taking these time blocks as separate units, two-fold FCS maximizes the data use (less information loss) and increases the plausibility of MAR assumption [64]. When compared to "basic" MICE, two-fold FCS increased precision and although some bias can still be present in the coefficients, it can be reduced by combining it with predictive mean matching (PMM) in combination with two-fold FCS [53].

Likelihood-based Methods

Two different likelihood-based methods for dealing with missing observations in single-level data were identified. Of them, one can be considered a full information maximum likelihood method: model-based methods using maximum likelihood via Monte Carlo Expectation Maximization (MCEM) (2/42 studies); and one a restricted maximum likelihood method (REML): sample selection models (3/42 studies). Such maximum likelihood-based methods are characterized by explicitly modelling the likelihood of having complete data (i.e., auxiliary missingness model), and simultaneously implementing the analysis model [65]. A disadvantage of maximum likelihood methods is that they are generally restricted to linear regression models [65].

Sample selection models can be used as a sensitivity analysis when MNAR is suspected, to explore possible departures from MAR [66, 67, 68]. Sample selection models (SMs), also known as **Heckman's model (HM)** or **Tobit type-2 model**, were originally developed to deal with selection bias. SMs model the mechanism by which the data are observed (i.e., selected) and is specified as a function of the underlying values [69]. SMs estimate two analysis models: one model for the outcome (i.e., equation of the outcome of interest), and a second model for the missing data mechanism (i.e., selection equation) [67, 68, 70, 71]. The outcome and selection equations are joined via their error terms through a bivariate normal distribution [67, 68, 70, 71]. As a result, SMs are primarily applied to missing outcomes. SMs can be used both for both continuous and binary outcomes. For continuous outcomes, two variations have been proposed: a one-step maximum likelihood estimator, and two-step maximum likelihood estimator [67]. For binary outcomes, a bivariate probit model and a one-step maximum likelihood estimator have been proposed [67]. Compared to CCA and MICE, when the outcome missing values were assumed to be MNAR, sample selection models result in less biased estimates [71].

The main limitation when using SMs, as previously stated, is that they are suitable to deal for handling missing outcome data. An extension of HM within the multiple imputation framework using a one or two-step maximum estimation approach,

namely multiple imputation using Heckman's model estimation, has been proposed to deal with missing observations in both outcomes and covariates [54, 67, 70]. Under both MAR and MNAR mechanisms, multiple imputation using Heckman's model estimation resulted in unbiased estimates of the regression coefficients [67, 70]. When compared with standard multiple imputation, it results in less biased estimates and coverage were closer to nominal value. However, when the Heckman's model is not fully compatible with the MNAR mechanism (i.e., non-Heckman scenario), the root mean square error (RMSE) (i.e., measures differences between values predicted by a model and the observed values) of the estimates is increased [70]. Galimard et al. noted that although multiple imputation using Heckman's model performs well when missing observations are MAR, researchers should always determine whether the data are most likely to have a MAR or MNAR missing data mechanism in order to avoid modelling a selection equation [67]. This can be done using a sensitivity analysis to explore departures from MNAR [67]. Other explored selection models were maximum likelihood selection model based on MAR mechanism [54, 68] and the Diggle-Kenward (DK) selection model when a MNAR mechanism is suspected [68].

Weighting Methods

One weighting method (WM) for dealing with missing observations in single-level data was identified: inverse probability weighting (7/42 studies).

Inverse probability weighting (IPW) estimates the likelihood of having complete data given the observed data (generally via logistic regression) and then uses the inverse of the probability of being a complete case to assign a weight to the observed cases in a weighted CCA [65, 66, 72]. In contrast to maximum-likelihood based methods, IPW does not model the distribution of variables with missing observations, but of the missingness mechanism [65, 66, 73]. When data is MAR, IPW results in unbiased estimates if the missingness models are correctly specified, because IPW is highly sensitive to model misspecification [74]. When dealing with MNAR outcome data, IPW resulted in substantial levels of bias [66]. When dealing when MAR, in comparison with multiple imputation and full information maximum likelihood based-methods (FIMLs), the RMSEs of scenarios using IPW were typically too large to make IPW the most valid strategy [66]. Another important methodological aspect regarding IPW is that the use of weights introduces a design effect similar to clustering (i.e., certain groups of individuals are likely to have similar weights), which increases standard errors and reduces statistical power [66]. However, when applying these weights to non-monotone missing observations, which is regularly the case in real-world data (e.g., an individual is discharged at some point in time, but is referred back later in time due to worsening of their condition), the weighting might result in a substantial loss of information because it is restricted to fully observed variables [65, 74]. When cases have occasional missing observations, a **double robust IPW (IPW-DR)** has been proposed [74, 75]. In IPW-DR, two terms are added in the weighting estimating equation. One term for the complete cases (i.e., fully observed) and a second term based on both fully and partially observed cases conditional on observed data [75]. Then, a working regression model is fitted in order to predict the missing observations (i.e., covariates)

[75]. When data are MAR and at least one of the two working regression models is correctly specified, IPW-DR produces regression estimates with a level of bias that is similar compared to multiple imputation [75]. However, multiple imputation produces estimates with less uncertainty and that are more robust to distributional assumptions regarding the variables with missing observations than IPW-DR [75].

3.2 Identified Statistical Methods for Dealing with Missing Data in Multilevel Data

In total, 11 studies were identified that evaluated five different categories of methods to account for missing observations in multilevel data (see Table 2 for the methods). These categories are: 1) deletion methods (5/11 studies, 45%), 2) single imputation methods (1/10 studies, 9%), 3) maximum likelihood-based methods (2/11 studies, 18%), 4) single-level multiple imputation methods (7/11 studies, 64%), and 5) multilevel multiple imputation methods (8/10 studies, 73%).

Deletion Methods

One deletion method for dealing with missing observations in multilevel data was identified: CCA (5/11 studies). In the included studies, CCA was mainly used as a reference method against which more advanced imputation methods were compared. CCA has been used to deal with missing observations in level-1 and level-2 covariates because it is easy to use [21]. As in single-level data, CCA results in unbiased estimates only if data are MCAR [32]. However, the associated reduction in precision can be even larger than in single-level data when dealing with missing observations in level-2 variables, because CCA then leads to the exclusion of a whole cluster resulting in selection at the cluster-level [21, 32].

Single Imputation Methods

Two different single imputation methods were assessed for dealing with missing observations in multilevel data: 1) last observation carried forward (2/11 studies), and 2) stochastic regression imputation (1/11 studies).

Last observation carried forward (LOCF) can be used reliably for variables that can validly be assumed to be constant over time (e.g., gender) [21]. However, the lack of inclusion of temporal trends in variables that are not constant over time results in severely biased estimates, even when data is MCAR [21, 40, 76]. Moreover, the failure to incorporate uncertainty around the imputed values results in reduced variability and overly precise results [21, 40, 76]. **Baseline observation carried forward (BOCF)** is a variation of LOCF, with the difference that it uses baseline observations to impute missing values making it more conservative than LOCF. However, BOCF has the same underlying assumptions as LOCF and therefore has similar problems [21].

Stochastic regression imputation is a stochastic variant of regression imputation that attempts to overcome the underestimation of variability associated with simple regression imputation by adding an extra error term to the regression model that reflects the imputation uncertainty. The extra error term adds random noise to the predictions, while preserving the regression weights and the

correlation structures [21]. However, the method does not adjust standard errors for imputation uncertainty [36]. The Bayesian counterpart of stochastic regression imputation accounts for uncertainty not only by adding an extra error term, but by also estimating a distribution for the regression coefficient [22]. The Bayesian counterpart of stochastic regression results in more valid estimates of uncertainty in the imputations compared to Frequentist stochastic regression [21].

Likelihood-based Methods

Two likelihood-based methods for dealing with missing observations in multilevel data were identified: 1) mixed-effects models (1/11 studies), and 2) pattern mixture models (1/11 studies). In the included studies, mixed-effects models were used as a method against which more advanced imputation methods were compared.

Mixed-effects models are an extension of linear regression models that incorporate both fixed-effects and random-effects parameters. Mixed-effect models are used in data that is clustered as well as with multiple assessments of the outcome (e.g., longitudinal data). Linear mixed-effect models use the full data set (i.e., complete and incomplete cases) and take into account the clusters in the data through a random effect. If data are MAR and the model is correctly specified, mixed-effect models will result in unbiased regression coefficients [21, 31, 77]. However, if the multilevel structure of the data is not taken into account in the imputation procedure, this it can result in biased estimates [77, 31]. In multilevel data, mixed-models can be fitted using likelihood-based methods to impute values in the outcome [21]. Additionally, mixed-effect models can be extended to not only include analysis variables in the model but also other covariates (i.e., auxiliary variables) [21, 32]. For example, this can be done by specifying a set of latent variables (i.e.,a random variable that is not observed or measured, but that can be inferred/estimated from the observed data using models) for the covariates with missing observations [32]. Also, **Bayesian analysis approaches** have been proposed to deal with missing observations in covariates [33, 78]. Goldstein et al.concluded that because information from all observations with missing data are included, the proposed Bayesian method resulted in smaller standard errors compared to standard multiple imputation [33]. However, Enders et al. indicated that in the currently available software, it might be challenging to implement the Bayesian methods due to the imposition of additional distributions in multilevel data (i.e., by extending a model so that the likelihood function will incorporate all variables with missing observations) [79].

Pattern mixture models (PMs) can be used as a sensitivity analysis when MNAR is suspected [53]. PMs formulate MNAR through the difference in distributions between the missing and observed values [69]. Observations are grouped based on their missing data pattern, and per missing data pattern a different model is specified [69, 66, 80]. A weighted sum of the conditional distribution of each missing data pattern is used to estimate the marginal distribution of the outcome [80]. PMs start by assuming an ignorable missing data mechanism (i.e., MAR or MCAR) and then incorporate the MNAR assumption into the model [53]. When a PM is combined with multiple imputation, first an ignorable missing data mechanism (i.e.

MAR or MCAR) is assumed and multiple imputation is used to impute the missing observations, resulting in multiple data sets. Next, PM is used and imputed values are adjusted to a fixed value. Larger adjustments of the imputed values indicate greater violation of the MAR or MCAR assumption. Finally, an analysis model (e.g., mixed-effects model) is fitted to and run for each imputed dataset after which the effect estimates are combined using Rubin's rules [53]. Compared to MICE, PMs may potentially reduce bias when informative attrition is suspected (i.e., when the reason for attrition is related the outcome which results in data being MNAR), and repeated outcome measurements are moderately correlated [53]. Similar to multiple imputation, including auxiliary variables in the imputation model may reduce residual bias [53]. However, there is uncertainty because there was only one study that evaluated PMs.

Single-level Multiple Imputation Methods

Two different single-level multiple imputation methods for dealing with missing observations in multilevel data were identified: 1) multiple imputation dummy indicator approach (DI) (1/11 studies), and 2) single-level multiple imputation by chained equations (MICE) (7/11 studies). Single-level multiple imputation methods, however, do not take the clustering of the data into account, which will lead to an underestimation of the intraclass correlations (ICC) [21].

Multiple imputation dummy indicator approach (MI-DI). The multiple imputation dummy indicator (MI-DI) approach can be considered a fixed-effects multiple imputation approach. MI-DI is an ad-hoc solution in which the multilevel structure of the data is taken into account by adding dummy indicator variables (i.e., N groups are represented by N-1 dummy variables or N indicator variables when the overall intercept is excluded) to the single-level multiple imputation model, which represent each cluster that the individual belongs to and to preserve the cluster level-variation [21, 77]. However, the MI-DI approach can result in overestimated regression coefficients and variance estimates, because it estimates separate intercepts for each cluster, which artificially inflates the true between cluster variation [21, 77]. In situations with large clusters and/or intraclass correlation coefficients (>0.30), bias was found to approach zero [77]. The MI-DI approach may lead to convergence problems of the imputation model, because a large number of variables needs to be included. In addition, the MI-DI approach is not suitable when missing observations are present in covariates [77, 81].

Multilevel Multiple Imputation Methods

Two different multiple imputation multilevel methods for dealing with missing observations in multilevel data were identified: 1) multilevel joint modeling (ML-JM) (3/11 studies), and 2) multilevel multiple imputation by chained equations (ML-MICE) (8/11 studies).

ML-JM and **MI-MICE** are multilevel extensions of single-level multiple imputation and are the most frequently used multiple imputation frameworks for multilevel data [34, 79, 81]. With increasing cluster sizes and differences between clusters

(which can lead to higher intercluster correlations), the multilevel structure needs to be accounted for in the imputation models to prevent biased estimates and overestimation of precision at group level [81, 82].The multilevel structure can be taken into account by modeling the multilevel structure using random-effects [1] (e.g., random-intercepts and slopes), or modeling relations between variables within and between groups (e.g., cross-level interactions) [32]. To obtain unbiased results in multilevel data, however, the imputation model should reflect the analysis model as closely as possible. Thus, the imputation model should preserve all the features and associations that are present in the analysis model [81].

Several multilevel data studies (n= 4) included in this review evaluated the performance of random-intercept models (i.e., different intercepts for each individual and for every individual the same slope; each individual has an exchangeable correlation coefficient) [32, 34, 81, 83]. ML-JM is similar to its single-level counterpart, in the sense that a joint imputation model is specified that uses a multivariate multilevel regression model to regress incomplete variables on complete variables [79]. ML-MICE on the other hand, imputes variables one by one from a sequence of univariate models with an incomplete variable regressed on complete and previously imputed variables [79]. The fact that previously imputed variables are also used as predictors is a crucial difference with JM [79].

There are also differences between ML-JM and ML-MICE in how they allow for the different relationships between variables to be modeled at individual-level and at group-level [32]. ML-JM allows relations to be different between hierarchical levels [81]. That is, ML-JM can be used to model different within- and between-cluster relations by means of a random-effect [32, 81], whereas ML-MICE assumes common associations between variables at level-1 and level-2 [81]. However, both methods assume similar covariance structures at individual and group levels [32]. Both ML-JM and ML-MICE have difficulty to preserve the variance associated with the random slope (i.e., random slope variation), although the effect of this differs between the methods. In the case of ML-JM it led to an overestimation of the slope variance, while ML-MICE led to an underestimation [79]. Another difference is that ML-JM treats all variables as outcomes without taking into account the missing data pattern [79, 81]. As a result, ML-JM and ML-MICE have a similar level of performance in terms of bias if only the outcome is missing and covariates are complete [79]. Moreover both ML-JM and ML-MICE resulted in an acceptable level of bias when there was 10% or less missing observations and the sample consisted of at least 30 clusters with 15 observations per group [79]. With increasing within-cluster sample sizes, the performance of the two methods was found to improve even more [79]. However, ML-JM performed somewhat better than ML-MICE when the intraclass correlation was high (e.g., ICC >= 0.50) [79], whereas ML-MICE appeared to have better performance than ML-JM when the intraclass correlation was low (e.g., ICC = 0.10) [79].

[1]If observations have a random-effect in common, they will be correlated and therefore dependencies in the data would have been modeled.

Although several studies focused on random-intercept models, random-coefficient models are also a possibility. However, only one study evaluated the performance of random-coefficient models (i.e., same intercepts, different slopes) [79] and showed that random-coefficient models resulted in diverse incompatibilities that can lead to bias. Full Bayesian imputation was suggested as a promising alternative in these scenarios [79].

4 Discussion

4.1 Main findings

This scoping review aimed to identify and critically appraise available statistical methods for dealing with missing data in real-world data studies. This review shows that a substantial amount of research has been performed in this area and that a wide range of promising methods is available for dealing with missing observations in real-world data (see Figure 2 for an overview of the methods). The scoping review's results confirm previous findings that relatively simple methods, such as CCA and single imputation, result in biased parameter estimates and a reduction of power, and are therefore not recommended when dealing with missing observations in real-world data. When missing observations are MAR, MICE and MVNI seem to perform best when imputing single-level data. Both methods deal with missing observations by creating a separate imputation and analysis model, while simultaneously taking imputation uncertainty into account by creating multiple imputed datasets. Moreover, both methods are relatively flexible in terms of their implementation. However, a disadvantage of both methods is that their standard specification does not account for non-normal distributions. Here, MICE has the advantage that several variations are available that can deal with non-normal data, whereas this is not the case for MVNI. For example, both MICE-PMM and MICE-CART can be used when data have non-normal distributions. Furthermore, temporal trends in time-varying covariates can be more easily taken into account with MICE than MVNI by using two-fold FCS.

When dealing with missing observations in multilevel data, it is important to take into account the multilevel structure (i.e., clustering) of the data, because the use of single-level methods can result in biased estimates when applied to clustered data [77, 81]. Our review shows that extensions of single-level methods multiple imputation methods (i.e., MICE and JM), which do consider the multilevel structure of the data through random-effects are the most frequently used and most flexible methods when dealing with MAR multilevel data. Mixed-effect models that combine imputation and analysis into one model are also used. However, implementation of mixed-effects models can be challenging due to the existence of multiple different distributions in multilevel data. Therefore, the methodological focus has mainly shifted towards multiple imputations-based multilevel strategies. However, it is important to take into account that the total number of included multilevel studies was limited and that the majority of these studies described very specific scenarios, thus generalizability of the results is not always possible.

Until now, most research has focused on MAR. However, in real-world data there

is also a high probability of missing observations being MNAR. Therefore, it is important to perform sensitivity analyses to explore the robustness of results to departures from MAR to MNAR. When dealing with MNAR data the list of available methods is less exhaustive, with pattern-mixture models and sample selection methods being the most researched [68]. For MAR methods mostly separate analysis and imputation models are being used, while MNAR methods have a combined model [68].

4.2 Comparison with existing research

Several reviews on handling missing observations in epidemiological studies have been previously published. However, the majority of these reviews focused on missing outcome data in RCTs [15, 17, 19], of which one focused specifically on clustered RCT data [16]. In real-world data missing observations often are present in both outcome and covariates or predictor variables, because data is not collected for research purposes but to support the healthcare process. Therefore, we included studies that dealt with missing observations in outcomes and/or covariates. Only one previous review by Karahalios et al. (2012) also focused on missing data approaches in observational data [20]. However, because of the rapid developments in methods that deal with missing observations over the last 10 years, there is little overlap in methods between the two reviews. To illustrate, compared to Karahalios et al. (2012), our review included more studies on advanced approaches, such as multiple imputation and full Bayesian modelling. Additionally, Karahalios et al. (2012) mainly included empirical studies [20], whereas we focused on simulation studies in which the performance of the methods can be formally assessed, because the "true vale" is known. In line with most reviews, we would like to stress the importance of sensitivity analyses regarding missing data mechanisms because using MAR techniques when data is actually MNAR will invariably result in biased estimates. Finally, in our review we looked separately at methods for single-level and multilevel data, whereas previous reviews did not make that distinction.

4.3 Strengths and limitations

To our knowledge, this is the first scoping review that summarizes the characteristics, advantages and disadvantages of methods that deal with missing data in real-world data. Another strength of our review is that it only includes simulation studies. Simulation studies allow us to understand the behavior of statistical methods, because the "true value" of a parameter of interest is known from the data generating mechanisms. This allows for the evaluation of the performance of the methods (e.g., bias, coverage, RMSE) [30]. In addition, we describe the specific circumstances under which the different methods were evaluated. For example, we provide information on the sample size, number of simulated data sets, types of missing variables (i.e., missing observations in covariates and/or outcome variables), percentage of missing data, assumed missing data mechanisms (i.e., MCAR, MAR, MNAR), and presence of hierarchical levels in the data. Finally, even though studies that compared only 'simple methods' were excluded, they are still portrayed in the categorization of the methods in order to present a complete panorama of available methods. However, we also need to acknowledge several of limitations.

First, we only searched for relevant studies using PubMed. We expect, however, that we have identified the great majority of methodological studies on the use of imputation methods in biomedical research in our review. Additionally, as with every review, important studies may have been missed [84]. However, this risk was minimized by using a broad search strategy including both Medical Subject Headings (MeSH) and a large set of specific keywords (i.e., descriptors and synonyms). Also, we searched reference lists of both relevant published reviews and the included studies to identify publications that we missed. Second, there is no critical appraisal tool available to assess the reliability and validity of the included simulation studies. However, by giving a detailed overview of the scenarios that were evaluated in the included simulation studies some comparison was possible. Third, only specific scenarios are evaluated in the simulation studies. Although, one can argue that this limitation applies to an even bigger extent to empirical studies, it is important that researchers check whether the scenarios evaluated in the simulation studies are representative of their own study. Third, the majority of the included studies explored the performance of multiple imputation-based methods. However, we complemented the overview of methods that had a few included studies with additional literature (e.g., by complementing the description of the methods and well known performance characteristics using textbooks). Finally, several variations of the discussed methods are available which not all have been discussed into detail. Instead, we provide a description of the most frequently found variants of the discussed methods, while a detailed overview of the variations of the included methods with their respective references in Supplementary Table 1.

4.4 Implications for further research and practice

The large number of identified statistical methods shows that researchers have a large arsenal of methods to choose from when dealing with missing data in real-world data-based studies. Which specific imputation method is most valid depends on a number of different aspects. Aspects that need to be taken into consideration are the distribution of the variables with missing observations, the presence of hierarchical levels, whether observations are missing in covariates, outcomes, or both, and what the most plausible missing data mechanism is. The results from the current review which are summarized in Figure 2, can be used as a reference to support researchers choosing a specific method. Additionally, many of the included studies recommended to assess the robustness of the results by assuming different missing data mechanisms, because using techniques that assume MAR when data is actually MNAR will result in bias. However, the number of simulation studies evaluating methods for MNAR data is limited. Thus, this area needs further research.

When performing observational studies, researchers are recommended to follow STROBE (i.e., STrengthening the Reporting of OBservational studies in Epidemiology) guidelines [85]. The STROBE guidelines recommend researchers to explain how missing data were addressed. However, they do not provide guidance on which statistical methods to use for dealing with missing data, even though they are inherent methodological issue of such studies. Additionally, the STROBE guidelines

do not contain a specific recommendation on the performance of sensitivity analyses to assess the robustness of assumptions about the missing data mechanism. Therefore, we recommend researchers to follow Sterne et al.(2009) who published more detailed guidelines for reporting analyses affected by missing observations [86]. Our review complements these guidelines by providing more detailed information on specific methods to impute missing observations.

5 Conclusion

This is the first scoping review that has described and critically appraised available statistical methods for dealing with missing observations in real-world data. Several aspects need to be considered when selecting a method: the types of missing variables (i.e., outcome and/or covariates), the distribution of the data, the presence of hierarchy levels in the data (i.e., single-level or multilevel), and assumed missing data mechanisms (i.e., MCAR, MAR, MNAR). Although MAR is the most commonly assumed missing data mechanism, data that is MNAR is also common in real-world data. Therefore, researchers are advised to perform sensitivity analyses to check the robustness of their data to the missing data mechanism assumptions. Various methods are available when MAR is assumed of which multiple imputation-based methods have the best performance. Therefore, we recommend the use of these methods when dealing with MAR observations in real-world data.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Availability of data and material
Not applicable, since this is a scoping literature review.

Competing interests
The authors have no competing interests to declare that are relevant to the content of this article.

Funding
Funding was provided by Netherlands Organisation for Scientific Research/ The Netherlands Organisation for Health Research and Development (NWO/ZonMw), (project ID no. 91717368; www.zonmw.nl/en/). NWO/ZonMw had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author's contributions
AEGM, ANV, MWH, JMvD, JFvS, MWvT, and JEB have contributed to the design of the study. AEGM and ANV screened studies for inclusion. AEGM extracted data from the included studies. ANV and JMvD assessed the data extraction. Data synthesis was performed by AEGM and supervised by JEB and JMvD. The first draft of the manuscript was written by AEGM. All authors read and approved the final manuscript.

Acknowledgements
Not applicable.

Author details
[1] Department of Health Sciences, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam Public Health Research Institute, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. [2]Department of Epidemiology and Data Science, Amsterdam University Medical Center, Amsterdam, The Netherlands. [3]Department of Health Sciences, Faculty of Science, Vrije Universiteit Amsterdam, MOVE Research Institute, Amsterdam, The Netherlands. [4]Amsterdam UMC location Vrije Universiteit Amsterdam, Department of Psychiatry, Amsterdam, The Netherlands. [5]Amsterdam Public Health, Mental Health program, Amsterdam, The Netherlands. [6]Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam Movement Science Research Institute, Amsterdam, The Netherlands.

References

1. Nick Freemantle and Thomas Strack. Real-world effectiveness of new medicines should be evaluated by appropriately designed clinical trials. *Journal of clinical epidemiology*, 63(10):1053–1058, 2010.
2. Victoria L Bartlett, Sanket S Dhruva, Nilay D Shah, Patrick Ryan, and Joseph S Ross. Feasibility of using real-world data to replicate clinical trial evidence. *JAMA network open*, 2(10):e1912869–e1912869, 2019.
3. Rachel E Sherman, Steven A Anderson, Gerald J Dal Pan, Gerry W Gray, Thomas Gross, Nina L Hunter, Lisa LaVange, Danica Marinac-Dabic, Peter W Marks, Melissa A Robb, et al. Real-world evidence—what is it and what can it tell us. *N Engl J Med*, 375(23):2293–2297, 2016.
4. Lawrence Blonde, Kamlesh Khunti, Stewart B Harris, Casey Meizinger, and Neil S Skolnik. Interpretation and impact of real-world clinical data for the practicing clinician. *Advances in therapy*, 35(11):1763–1774, 2018.
5. Elodie Baumfeld Andre, Robert Reynolds, Patrick Caubel, Laurent Azoulay, and Nancy A Dreyer. Trial designs using real-world data: The changing landscape of the regulatory approval process. *Pharmacoepidemiology and Drug Safety*, 2019.
6. Amr Makady, Anthonius de Boer, Hans Hillege, Olaf Klungel, Wim Goettsch, et al. What is real-world data? a review of definitions based on literature and stakeholder interviews. *Value in health*, 20(7):858–865, 2017.
7. Lars G Hemkens, Despina G Contopoulos-Ioannidis, and John PA Ioannidis. Current use of routinely collected health data to complement randomized controlled trials: a meta-epidemiological survey. *CMAJ open*, 4(2):E132, 2016.
8. Jonathan P Jarow, Lisa LaVange, and Janet Woodcock. Multidimensional evidence generation and fda regulatory decision making: defining and using "real-world" data. *Jama*, 318(8):703–704, 2017.
9. Sophie Graham, Laura McDonald, Radek Wasiak, Michael Lees, and Sreeram Ramagopalan. Time to really share real-world data? *F1000Research*, 7, 2018.
10. Sheldon Greenfield. Making real-world evidence more useful for decision making. *Value in Health*, 20(8):1023–1024, 2017.
11. Eric I Benchimol, Liam Smeeth, Astrid Guttmann, Katie Harron, David Moher, Irene Petersen, Henrik T Sørensen, Erik von Elm, Sinéad M Langan, RECORD Working Committee, et al. The reporting of studies conducted using observational routinely-collected health data (record) statement. *PLoS Med*, 12(10):e1001885, 2015.
12. Wendy Lipworth. Real-world data to generate evidence about healthcare interventions. *Asian Bioethics Review*, 11(3):289–298, 2019.
13. Brian J Wells, Kevin M Chagin, Amy S Nowacki, and Michael W Kattan. Strategies for handling missing data in electronic health record derived data. *Egems*, 1(3), 2013.
14. Eric I Benchimol, Liam Smeeth, Astrid Guttmann, Katie Harron, David Moher, Irene Petersen, Henrik T Sørensen, Erik von Elm, Sinéad M Langan, RECORD Working Committee, et al. The reporting of studies conducted using observational routinely-collected health data (record) statement. *PLoS Med*, 12(10):e1001885, 2015.
15. Melanie L Bell, Mallorie Fiero, Nicholas J Horton, and Chiu-Hsieh Hsu. Handling missing data in rcts; a review of the top medical journals. *BMC medical research methodology*, 14(1):118, 2014.

16. Karla Díaz-Ordaz, Michael G Kenward, Abie Cohen, Claire L Coleman, and Sandra Eldridge. Are missing data adequately handled in cluster randomised trials? a systematic review and guidelines. *Clinical Trials*, 11(5):590–600, 2014.
17. Thomas R Sullivan, Lisa N Yelland, Katherine J Lee, Philip Ryan, and Amy B Salter. Treatment of missing data in follow-up studies of randomised controlled trials: A systematic review of the literature. *Clinical Trials*, 14(4):387–395, 2017.
18. Iris Eekhout, Michiel R. de Boer, Jos W. R. Twisk, Henrica C. W. de Vet, and Martijn W. Heymans. Missing data a systematic review of how they are reported and handled. *Epidemiology*, 23(5):729–732, 9 2012.
19. Angela M Wood, Ian R White, and Simon G Thompson. Are missing outcome data adequately handled? a review of published randomized controlled trials in major medical journals. *Clinical trials*, 1(4):368–376, 2004.
20. Amalia Karahalios, Laura Baglietto, John B Carlin, Dallas R English, and Julie A Simpson. A review of the reporting and handling of missing data in cohort studies with repeated assessment of exposure measures. *BMC medical research methodology*, 12(1):96, 2012.
21. Stef van Buuren. *Flexible Imputation of Missing Data*. Chapman & Hall / CRC, second edition, 2018.
22. Martijn W Heijmans and Iris Eekhout. *Applied Missing Data Analysis With SPSS and (R)Studio*. First edition, 2019.
23. Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*, volume 793. John Wiley & Sons, 2019.
24. Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*, volume 793. John Wiley & Sons, 2019.
25. Donald B Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.
26. Evangelos Kontopantelis, Ian R White, Matthew Sperrin, and Iain Buchan. Outcome-sensitive multiple imputation: a simulation study. *BMC medical research methodology*, 17(1):2, 2017.
27. Thomas R Sullivan, Ian R White, Amy B Salter, Philip Ryan, and Katherine J Lee. Should multiple imputation be the method of choice for handling missing data in randomized trials? *Statistical methods in medical research*, 27(9):2610–2626, 2018.
28. Krishnan Bhaskaran and Liam Smeeth. What is the difference between missing completely at random and missing at random? *International journal of epidemiology*, 43(4):1336–1339, 2014.
29. Jonathan AC Sterne, Ian R White, John B Carlin, Michael Spratt, Patrick Royston, Michael G Kenward, Angela M Wood, and James R Carpenter. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj*, 338:b2393, 2009.
30. Tim P Morris, Ian R White, and Michael J Crowther. Using simulation studies to evaluate statistical methods. *Statistics in medicine*, 38(11):2074–2102, 2019.
31. Vincent Audigier, Ian R. White, Shahab Jolani, Thomas P. A. Debray, Matteo Quartagno, James Carpenter, Stef van Buuren, and Matthieu Resche-Rigon. Multiple imputation for multilevel data with continuous and binary variables. *Statistical Science*, 33(2):160–183, 5 2018.
32. Simon Grund, Oliver Lüdtke, and Alexander Robitzsch. Multiple imputation of missing data for multilevel models: Simulations and recommendations. *Organizational Research Methods*, 21(1):111–149, 2018.
33. Harvey Goldstein, James R Carpenter, and William J Browne. Fitting multilevel multivariate models with missing data in responses and covariates that may include interactions and non-linear terms. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pages 553–564, 2014.
34. Craig K Enders, Brian T Keller, and Roy Levy. A fully conditional specification approach to multilevel imputation of categorical and continuous variables. *Psychological methods*, 23(2):298, 2018.
35. Therese D Pigott. A review of methods for missing data. *Educational research and evaluation*, 7(4):353–383, 2001.
36. Amanda N. Baraldi and Craig K. Enders. An introduction to modern missing data analyses. *Journal of School Psychology*, 48(1):5 – 37, 2010.
37. Jonathan W. Bartlett, James R. Carpenter, Kate Tilling, and Stijn Vansteelandt. Improving upon the efficiency of complete case analysis when covariates are MNAR. *Biostatistics*, 15(4):719–730, 06 2014.
38. Rachael A Hughes, Jon Heron, Jonathan A C Sterne, and Kate Tilling. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *International Journal of Epidemiology*, 48(4):1294–1304, 03 2019.
39. Rolf HH Groenwold, A Rogier T Donders, Kit CB Roes, Frank E Harrell Jr, and Karel GM Moons. Dealing with missing outcome data in randomized trials and observational studies. *American journal of epidemiology*, 175(3):210–217, 2012.
40. Mary B. Roberts, Mary C. Sullivan, and Suzy B. Winchester. Examining solutions to missing data in longitudinal nursing research. *Journal for Specialists in Pediatric Nursing*, 22(2):e12179, 2017.
41. Gabriel L. Schlomer, Sheri Bauman, and Noel A. Card. Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology*, 57(1):1 – 10, 2010.
42. Jeanne A. Saunders, Nancy Morrow-Howell, Edward Spitznagel, Peter Doré, Enola K. Proctor, and Richard Pescarino. Imputing Missing Data: A Comparison of Methods for Social Work Researchers. *Social Work Research*, 30(1):19–31, 03 2006.
43. Yulei He, Recai M. Yucel, and Trivellore E. Raghunathan. A functional multiple imputation approach to incomplete longitudinal data. *Statistics in medicine*, 30 10:1137–56, 2011.
44. Katherine J Lee and John B Carlin. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *American journal of epidemiology*, 171(5):624–632, 2010.
45. Katherine J Lee and John B Carlin. Recovery of information from multiple imputation: a simulation study. *Emerging themes in epidemiology*, 9(1):1–10, 2012.
46. Anurika Priyanjali De Silva, Margarita Moreno-Betancur, Alysha Madhu De Livera, Katherine Jane Lee, and Julie Anne Simpson. A comparison of multiple imputation methods for handling missing values in longitudinal data in the presence of a time-varying covariate with a non-linear association with time: a simulation study. *BMC medical research methodology*, 17(1):1–11, 2017.

47. Anurika Priyanjali De Silva, Margarita Moreno-Betancur, Alysha Madhu De Livera, Katherine Jane Lee, and Julie Anne Simpson. Multiple imputation methods for handling missing values in a longitudinal categorical variable with restrictions on transitions over time: a simulation study. *BMC medical research methodology*, 19(1):1–14, 2019.
48. Md Hamidul Huque, John B Carlin, Julie A Simpson, and Katherine J Lee. A comparison of multiple imputation methods for missing data in longitudinal studies. *BMC medical research methodology*, 18(1):1–16, 2018.
49. Laura Rodwell, Katherine J Lee, Helena Romaniuk, and John B Carlin. Comparison of methods for imputing limited-range variables: a simulation study. *BMC medical research methodology*, 14(1):1–11, 2014.
50. Anne-Françoise Donneau, Murielle Mauer, Philippe Lambert, Geert Molenberghs, and Adelin Albert. Simulation-based study comparing multiple imputation methods for non-monotone missing ordinal data in longitudinal settings. *Journal of biopharmaceutical statistics*, 25(3):570–601, 2015.
51. Yusuke Yamaguchi, Toshihiro Misumi, and Kazushi Maruo. A comparison of multiple imputation methods for incomplete longitudinal binary data. *Journal of biopharmaceutical statistics*, 28(4):645–667, 2018.
52. Jonathan W Bartlett, Shaun R Seaman, Ian R White, James R Carpenter, and Alzheimer's Disease Neuroimaging Initiative*. Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. *Statistical methods in medical research*, 24(4):462–487, 2015.
53. Catherine A Welch, Séverine Sabia, Eric Brunner, Mika Kivimäki, and Martin J Shipley. Does pattern mixture modelling reduce bias due to informative attrition compared to fitting a mixed effects model to the available cases or data imputed using multiple imputation?: a simulation study. *BMC medical research methodology*, 18(1):1–15, 2018.
54. Emmanuel O Ogundimu and Gary S Collins. A robust imputation method for missing responses and covariates in sample selection models. *Statistical methods in medical research*, 28(1):102–116, 2019.
55. Stef Van Buuren. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical methods in medical research*, 16(3):219–242, 2007.
56. Nicole S Erler, Dimitris Rizopoulos, Joost van Rosmalen, Vincent WV Jaddoe, Oscar H Franco, and Emmanuel MEH Lesaffre. Dealing with missing covariates in epidemiologic studies: a comparison between multiple imputation and a full bayesian approach. *Statistics in medicine*, 35(17):2955–2974, 2016.
57. Hua Yun Chen, Hui Xie, and Yi Qian. Multiple imputation for missing values through conditional semiparametric odds ratio models. *Biometrics*, 67(3):799–809, 2011.
58. Kyle M Lang and Wei Wu. A comparison of methods for creating multiple imputations of nominal variables. *Multivariate behavioral research*, 52(3):290–304, 2017.
59. Katherine J Lee and John B Carlin. Multiple imputation in the presence of non-normal data. *Statistics in Medicine*, 36(4):606–617, 2017.
60. Tim P Morris, Ian R White, and Patrick Royston. Tuning multiple imputation by predictive mean matching and local residual draws. *BMC medical research methodology*, 14(1):1–13, 2014.
61. Dandan Xu, Michael J Daniels, and Almut G Winterstein. Sequential bart for imputation of missing covariates. *Biostatistics*, 17(3):589–602, 2016.
62. Chiu-Hsieh Hsu, Yulei He, Yisheng Li, Qi Long, and Randall Friese. Doubly robust multiple imputation using kernel-based techniques. *Biometrical Journal*, 58(3):588–606, 2016.
63. Lane F Burgette and Jerome P Reiter. Multiple imputation for missing data via sequential regression trees. *American journal of epidemiology*, 172(9):1070–1076, 2010.
64. Catherine A Welch, Irene Petersen, Jonathan W Bartlett, Ian R White, Louise Marston, Richard W Morris, Irwin Nazareth, Kate Walters, and James Carpenter. Evaluation of two-fold fully conditional specification multiple imputation for longitudinal electronic health record data. *Statistics in medicine*, 33(21):3725–3737, 2014.
65. James C Doidge. Responsiveness-informed multiple imputation and inverse probability-weighting in cohort studies with missing data that are non-monotone or not missing at random. *Statistical methods in medical research*, 27(2):352–363, 2018.
66. Rebekah Young and David Johnson. Methods for handling missing secondary respondent data. *Journal of Marriage and Family*, 75(1):221–234, 2013.
67. Jacques-Emmanuel Galimard, Sylvie Chevret, Emmanuel Curis, and Matthieu Resche-Rigon. Heckman imputation models for binary or continuous mnar outcomes and mar predictors. *BMC medical research methodology*, 18(1):1–13, 2018.
68. Meijuan Li, Nan Chen, Yang Cui, and Hongyun Liu. Comparison of different lgm-based methods with mar and mnar dropout data. *Frontiers in psychology*, 8:722, 2017.
69. Baptiste Leurent, Manuel Gomes, Rita Faria, Stephen Morris, Richard Grieve, and James R Carpenter. Sensitivity analysis for not-at-random missing data in trial-based cost-effectiveness analysis: a tutorial. *Pharmacoeconomics*, 36(8):889–901, 2018.
70. Jacques-Emmanuel Galimard, Sylvie Chevret, Camelia Protopopescu, and Matthieu Resche-Rigon. A multiple imputation approach for mnar mechanisms compatible with heckman's model. *Statistics in medicine*, 35(17):2907–2920, 2016.
71. Siaka Koné, Bassirou Bonfoh, Daouda Dao, Inza Koné, and Günther Fink. Heckman-type selection models to obtain unbiased estimates with missing measures outcome: theoretical considerations and an application to missing birth weight data. *BMC medical research methodology*, 19(1):1–13, 2019.
72. Shaun R Seaman and Ian R White. Review of inverse probability weighting for dealing with missing data. *Statistical methods in medical research*, 22(3):278–295, 2013.
73. Antoine Lewin, Ruben Brondeel, Tarik Benmarhnia, Frédérique Thomas, and Basile Chaix. Attrition bias related to missing outcome data. *Epidemiology*, 29(1):87–95, 2018.
74. Shaun R Seaman, Ian R White, Andrew J Copas, and Leah Li. Combining multiple imputation and inverse-probability weighting. *Biometrics*, 68(1):129–137, 2012.
75. Chiu-Hsieh Hsu, Qi Long, Yisheng Li, and Elizabeth Jacobs. A nonparametric multiple imputation approach

for data with missing covariate values with application to colorectal adenoma data. *Journal of biopharmaceutical statistics*, 24(3):634–648, 2014.
76. Rita Faria, Manuel Gomes, David Epstein, and Ian R. White. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. *PharmacoEconomics*, 32(12):1157–1170, Dec 2014.
77. Oliver Lüdtke, Alexander Robitzsch, and Simon Grund. Multiple imputation of missing data in multilevel designs: A comparison of different strategies. *Psychological Methods*, 22:141–165, 2017.
78. Nicole S Erler, Dimitris Rizopoulos, Joost van Rosmalen, Vincent WV Jaddoe, Oscar H Franco, and Emmanuel MEH Lesaffre. Dealing with missing covariates in epidemiologic studies: a comparison between multiple imputation and a full bayesian approach. *Statistics in medicine*, 35(17):2955–2974, 2016.
79. Craig K. Enders, Timothy Hayes, and Han Du. A comparison of multilevel imputation schemes for random coefficient models: Fully conditional specification and joint model imputation with random covariance matrices. *Multivariate behavioral research*, 53 5:695–713, 2018.
80. Changyu Shen and Lisa Weissfeld. Application of pattern-mixture models to outcomes that are potentially missing not at random using pseudo maximum likelihood estimation. *Biostatistics*, 6(2):333–347, 2005.
81. Craig K. Enders, Stephen A. Mistler, and Brian T. Keller. Multilevel multiple imputation: A review and evaluation of joint modeling and chained equations imputation. *Psychological Methods*, 12 2015.
82. Mallorie Fiero, Shuang Huang, and Melanie L Bell. Statistical analysis and handling of missing data in cluster randomised trials: protocol for a systematic review. *BMJ Open*, 5(5), 2015.
83. Nisha C Gottfredson, Sonya K Sterba, and Kristina M Jackson. Explicating the conditions under which multilevel multiple imputation mitigates bias resulting from random coefficient-dependent missing longitudinal data. *Prevention Science*, 18(1):12–19, 2017.
84. José Antonio Salvador-Oliván, Gonzalo Marco-Cuenca, and Rosario Arquero-Avilés. Errors in search strategies used in systematic reviews and their effects on information retrieval. *Journal of the Medical Library Association: JMLA*, 107(2):210, 2019.
85. Andre Knottnerus and Peter Tugwell. Strobe—a checklist to strengthen the reporting of observational studies in epidemiology. *Journal of clinical epidemiology*, 61(4):323, 2008.
86. Jonathan AC Sterne, Ian R White, John B Carlin, Michael Spratt, Patrick Royston, Michael G Kenward, Angela M Wood, and James R Carpenter. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj*, 338, 2009.
87. Tra My Pham, James R Carpenter, Tim P Morris, Angela M Wood, and Irene Petersen. Population-calibrated multiple imputation for a binary/categorical covariate in categorical regression models. *Statistics in medicine*, 38(5):792–808, 2019.
88. RP Cornish, John Macleod, JR Carpenter, and Kate Tilling. Multiple imputation using linked proxy outcome data resulted in important bias reduction and efficiency gains: a simulation study. *Emerging themes in epidemiology*, 14(1):1–13, 2017.
89. Ying Ding, Shengchun Kong, Shan Kang, and Wei Chen. A semiparametric imputation approach for regression with censored covariate with application to an amd progression study. *Statistics in medicine*, 37(23):3293–3308, 2018.
90. Mark A Ferro. Missing data in longitudinal studies: cross-sectional multiple imputation provides similar estimates to full-information maximum likelihood. *Annals of epidemiology*, 24(1):75–77, 2014.
91. Qingxia Chen and Joseph G Ibrahim. A note on the relationships between multiple imputation, maximum likelihood and fully bayesian methods for missing responses in linear regression models. *Statistics and its Interface*, 6(3):315, 2014.
92. Shahab Jolani, Laurence E Frank, and Stef van Buuren. Dual imputation model for incomplete longitudinal data. *British Journal of Mathematical and Statistical Psychology*, 67(2):197–212, 2014.
93. Hiroaki Tomita, Hironori Fujisawa, and Masayuki Henmi. A bias-corrected estimator in multiple imputation for missing data. *Statistics in Medicine*, 37(23):3373–3386, 2018.
94. Ian R White and John B Carlin. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Statistics in medicine*, 29(28):2920–2931, 2010.
95. Shaun R Seaman, Jonathan W Bartlett, and Ian R White. Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods. *BMC medical research methodology*, 12(1):1–13, 2012.
96. Shao-Hsien Liu, Stavroula A Chrysanthopoulou, Qiuzhi Chang, Jacob N Hunnicutt, and Kate L Lapane. Missing data in marginal structural models: a plasmode simulation study comparing multiple imputation and inverse probability weighting. *Medical care*, 57(3):237, 2019.
97. Jonathan W Bartlett, James R Carpenter, Kate Tilling, and Stijn Vansteelandt. Improving upon the efficiency of complete case analysis when covariates are mnar. *Biostatistics*, 15(4):719–730, 2014.
98. Muhan Zhou, Yulei He, Mandi Yu, and Chiu-Hsieh Hsu. A nonparametric multiple imputation approach for missing categorical data. *BMC medical research methodology*, 17(1):1–12, 2017.
99. Cuiling Wang and Charles B Hall. Correction of bias from non-random missing longitudinal data using auxiliary information. *Statistics in medicine*, 29(6):671–679, 2010.
100. Simon Grund, Oliver Lüdtke, and Alexander Robitzsch. Multiple imputation of missing covariate values in multilevel models with random slopes: a cautionary note. *Behavior Research Methods*, 48:640–649, 2016.
101. Hua Fang. Mifuzzy clustering for incomplete longitudinal data in smart health. *Smart Health*, 1:50–65, 2017.

Figures

Figure 1: **PRISMA flow chart of identification, screening, eligibility and inclusion of studies**

Figure 2: **Summary of identified methods for dealing with missing data per hierarchical level**

Table 1. Identified statistical methods dealing with missing data in single-level data

Category	Methods	Assumed Missing Data Mechanism	Available Software	Remarks	References
Simple Methods					
Deletion	Complete-case analysis (CCA), Complete records analysis (CRA), Listwise deletion	MCAR	SPSS, SAS, STATA, S-PLUS,	[1] Unbiased estimates only if data are MCAR and when only the outcome is missing.	[26, 43, 44, 45, 46, 47, 52, 57, 62, 64, 65, 66, 74, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98]
	Available case analysis (ACA), pairwise deletion	MCAR	SPSS, SAS, STATA, R	[1] Unbiased estimates only if data are MCAR.	[47, 53, 98]
Single imputation (SI)	Mean imputation, mean replacement	MCAR	SPSS, SAS, STATA, R	[1] Unbiased mean estimates only if the data are MCAR. [2] Variance is underestimated [3] Correlations are underestimated.	[40, 43]
	Regression imputation	MCAR	SPSS, SAS, STATA, R	[1] Unbiased mean estimates if data are MCAR. [2] Unbiased regression weight estimates if data are MCAR. [3] Unbiased estimates of regression weights if data are MAR only when all variables related to missingness are part of the regression model. [4] Variance is underestimated. [5] Correlations are overestimated.	[40]

Table 1. continued Identified statistical methods dealing with missing data in single-level data

Category	Methods	Assumed Missing Data Mechanism	Available Software	Remarks	References
Advanced Methods					
Multiple imputation (MI) methods	Multivariate imputation by chained equations (MICE), fully conditional specification (FCS), sequential regression imputation, ordered pseudo-Gibbs sampling, partially incompatible MCMC	MAR/MCAR/ MNAR	SPSS, SAS, IVEware (through SAS or standalone), STATA, R (mice package), S-plus, M-plus	[1] Unbiased regression estimates if the data are MAR. [2] Biased regression estimates when missing variables have skewed distributions. [3] Relative bias increases when percentage of missing data increases. [4] Unbiased estimates when missing values are continuous and/or categorical. [5] Does not take into account temporal trends at irregular time intervals.	[40, 44, 46, 47, 48, 49, 52, 53, 56, 57, 59, 61, 63, 64, 65, 96]
	Multivariate normal imputation (MVNI Deletion), multiple imputation based on multivariate normal distribution, joint modeling	MAR/MCAR	SPSS, SAS, STATA, R	[1] Unbiased mean estimates if the data are MAR. [2] Does not take into account temporal trends at irregular time intervals.	[44, 45, 46, 47, 48, 49, 57, 94, 98, 99]
Likelihood-based Methods	Model based method using maximum likelihood (ML) via Monte Carlo expectation maximization (MCEM)	MAR	SPSS, SAS, STATA, R	-	[90, 91]
	Selection models: sample selection model, Tobit type 2 model, Heckman's model	MNAR	SPSS, SAS, STATA, R	[1] Deals with missing outcomes in MNAR.	[54, 66, 68]
Weighting methods	Inverse probability weighting (IPW)	MAR	SPSS, SAS, STATA, R	[1] Unbiased regression estimated when data are MAR.	[65, 66, 74, 75, 96, 97, 98]

Table 2. Identified statistical methods dealing with missing data in multilevel data

Category	Methods	Assumed Missing Data Mechanism	Available Software	Remarks	References
Simple Methods					
Deletion	Complete case analysis (CCA), Complete records analysis (CRA), Listwise deletion	MCAR	SPSS, SAS, STATA, R	[1] Unbiased estimates only if data are MCAR and when only the outcome is missing. [2] Can provide reasonable estimates of the slope variance.	[31, 32, 33, 93, 100]
Single imputation	Last observation carried forward (LOCF), last value carried forward (LVCF) and baseline observation carried forward (BOCF)	MCAR	SPSS, SAS, STATA, R	[1] Does not adequately reflect the longitudinal nature of the data. [2] Variance is underestimated.	[40, 43]
	Mean imputation, mean replacement	MCAR	SPSS, SAS, STATA, R	[1] Does not adequately reflect the longitudinal nature of the data. [2] Variance is underestimated. [3] Does not adequately reflect the association between the explanatory variables.	[40]
	Stochastic regression imputation	MCAR, MAR	SPSS, SAS, STATA, R	[1] Unbiased regression coefficients. [2] Maximizes variability. [3] Unbiased variance estimates. [4] Variance is underestimated. [5] Correlations are preserved.	[40]
Advanced Methods					
Maximum-likelihood based methods	Pattern mixture models (PM)	MAR/ MNAR	SAS, STATA, R	[1] Can only be used to conduct sensitivity analysis to assess possible departures from MAR to MNAR.	[53]
	Mixed-Effects Models, linear mixed model	MAR	MPlus, R, WinBUGS , xxM, Latent GOLD	[1] Unbiased estimates when outcomes have missing. [2] Unbiased estimates of variance components values.	[32]

Table 2. continued Identified statistical methods dealing with missing data in multilevel data

Category	Methods	Assumed Missing Data Mechanism	Available Software	Remarks	References
Advanced Methods					
Single-level Multiple Imputation	Single-level multiple imputation by chained equations (SL-MICE)	MAR	R, SPSS, STATA	[1] Biased estimates of variance components. [2] Biased regression coefficients. [3] Negatively biased estimates of ICC (even with very large groups).	[32, 33, 77, 81, 83, 100, 101]
	Multiple imputation dummy indicator (DI) approach (including cluster variable)	MAR	R	[1] Biased estimates of variance components. [2] Biased regression coefficients. [3] Biased estimates when explanatory variables have missing values. [4] Positively biased estimates of the intraclass correlation coefficient (ICC).	[77]
Multilevel Multiple Imputation	Multilevel chained equations (MICE), multilevel fully conditional specification (FCS), two level model, two-level imputation	MAR	Mplus, R, Blimp	[1] Unbiased estimates when only outcomes have missing values. [2] Limited biased estimates when explanatory variables have missing values.	[31, 32, 34, 77, 79, 81, 83, 100]
	Joint Modeling, Joint Multilevel multiple imputation modeling approach, Multilevel Joint modeling (JM) , Multilevel joint modelling multiple imputation (JM-jomo)	MAR	Blimp, Mplus, MLwiN, R, Stata, SAS macro, Realcom-impute	[1] Biased estimates when the number of individuals and/or clusters is small. [2] Biased estimates when variance of random effects is small. [3] Biased estimates when explanatory variables have missing values. [4] Variance is overestimated. [5] Fails to preserve covariation between the intercepts and the slopes.	[31, 79, 81]

Figures

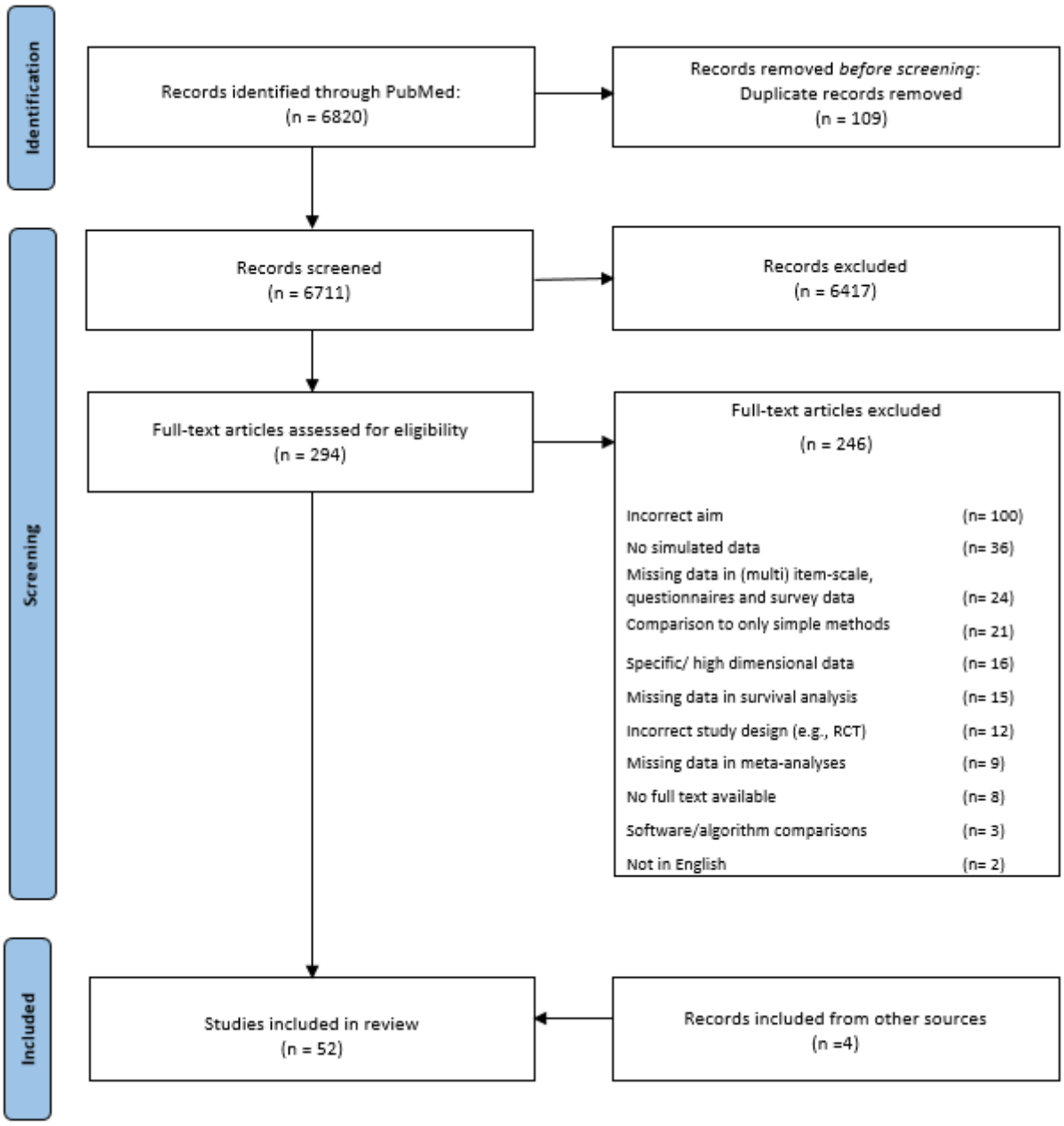

Figure 1

PRISMA flow chart of identification, screening, eligibility and inclusion of studies

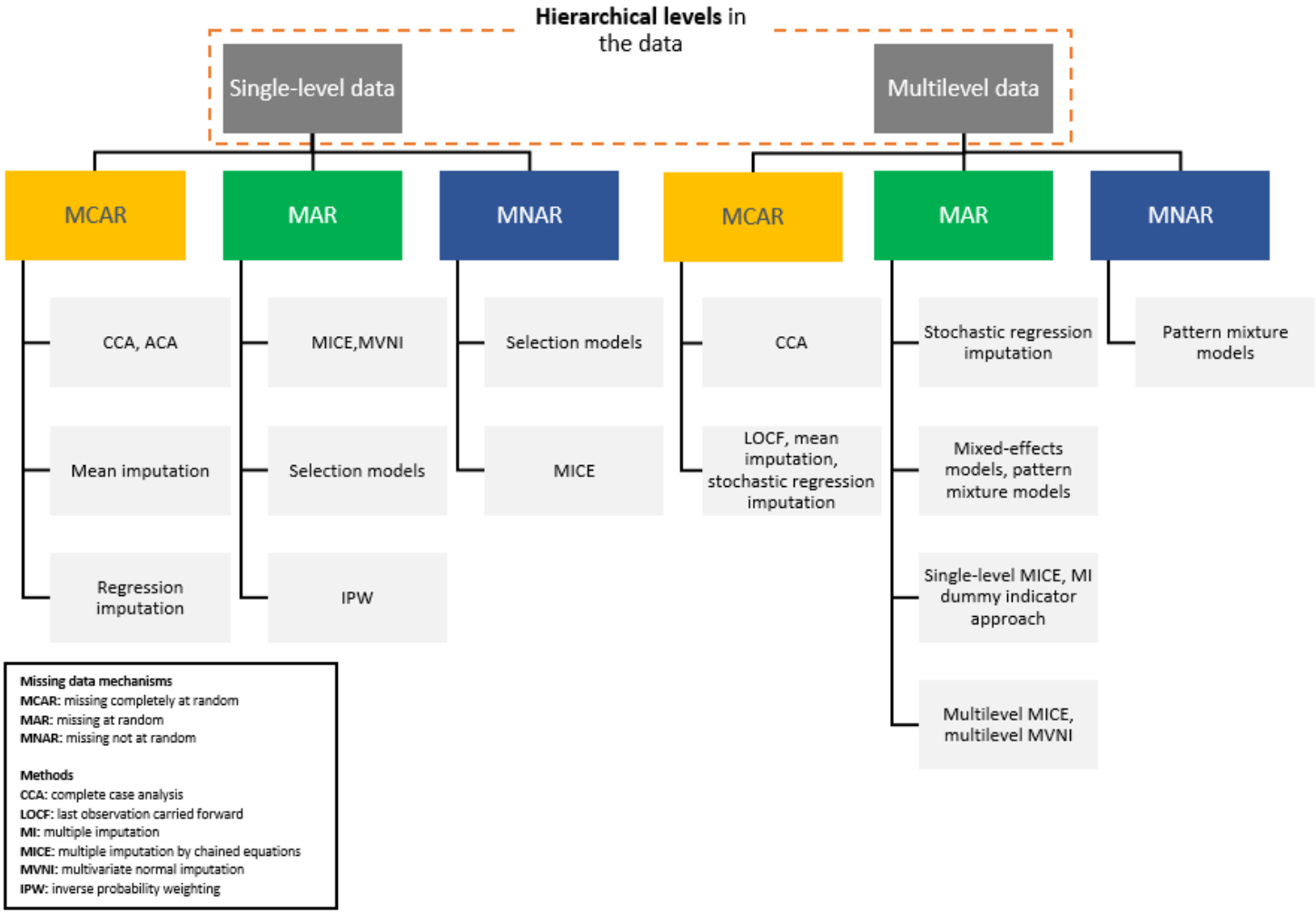

Figure 2

Summary of identified methods for dealing with missing data per hierarchical level

Supplementary Files

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

- Appendix1FullPubMedStrategy03052022.docx
- Appendix2ExamplesMissingDataPatterns03052022.docx
- Supplementary1Overviewincludedstudies03052022.docx
- Supplementary2SummaryDGM03052022.docx