

RESEARCH

Open Access



How much missing data is too much to impute for longitudinal health indicators? A preliminary guideline for the choice of the extent of missing proportion to impute with multiple imputation by chained equations

K. P. Junaid¹ , Tanvi Kiran^{1*} , Madhu Gupta¹ , Kamal Kishore² and Sujata Siwach³

Abstract

Background The multiple imputation by chained equations (MICE) is a widely used approach for handling missing data. However, its robustness, especially for high missing proportions in health indicators, is under-researched. The study aimed to provide a preliminary guideline for the choice of the extent of missing proportion to impute longitudinal health-related data using the MICE method.

Methods The study obtained complete data on five mortality-related health indicators of 100 countries (2015–2019) from the Global Health Observatory. Nine incomplete datasets with missing rates from 10 to 90% were generated and imputed using MICE. The robustness of MICE was assessed through three approaches: comparison of means using the Repeated Measures- Analysis of variance, estimation of evaluation metrics (Root mean square error, mean absolute deviation, Bias, and proportionate variance), and visual inspection of box plots of imputed and non-imputed data.

Results The Repeated Measures- Analysis of variance revealed significant differences between complete and imputed data, primarily in imputed data with over 50% missing proportions. Evaluation metrics exhibited 'high performance' for the dataset with a 50% missing proportion for various health indicators. However, with missing proportions exceeding 70%, the majority of indicators demonstrated a 'low' performance level in terms of most evaluation metrics. The visual inspection of the box plot revealed severe variance shrinkage in imputed datasets with missing proportions beyond 70%, corroborating the findings from the evaluation metrics.

Conclusion It demonstrates high robustness up to 50% missing values, with marginal deviations from complete datasets. Caution is warranted for missing proportions between 50 and 70%, as moderate alterations are observed. Proportions beyond 70% lead to significant variance shrinkage and compromised data reliability, emphasizing the importance of acknowledging imputation limitations for practical decision-making.

Keywords Multiple imputation, MICE, Missing proportion, RM-ANOVA, Mortality, Longitudinal data

*Correspondence:

Tanvi Kiran

tanvikiran3@yahoo.com

Full list of author information is available at the end of the article



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

Introduction

International organizations such as the World Health Organization (WHO) [1], United Nations Children's Fund (UNICEF) [2], and United Nations Program on HIV/AIDS (UNAIDS) [3] make available a vast amount of country-specific data on the enormous number of health indicators. For instance, The WHO's Global Health Observatory (GHO) provides access to health-related statistical data for 198 WHO member countries [1]. These data are often in the form of time-series data in order to monitor and track changes in the health status of countries. The global research community are utilizing these population health data across the country to undertake various analytical studies, such as interrupted and uninterrupted time-series analysis, geostatistical analyses, pre-post comparison, longitudinal analyses, and cross-sectional analyses [4]. Exploration of these population health data is crucial for efficient policy formulation and resource allocation. However, despite these databases providing a rich source of health-related information across different time points, researchers face challenges utilizing these country-specific population health data for research purposes mainly due to the issue of missing information [5]. Often longitudinal studies in health research encounter problems of missing information resulting in biasedness and less reliable estimates.

The missingness of population health data occurs due to multifaceted reasons. National health information systems across various countries often fail to produce reliable and complete data due to issues in population coverage, representativeness, frequency, timeliness, and disaggregation [6]. Missingness in national health datasets is exacerbated by poorly integrated data sources, lack of data-sharing standards, and inadequate skills among personnel handling data [7]. Resource limitations and political factors, including poor governance and instability, hinder the collection and reporting of accurate health data in many Low and Middle Income Countries (LMICs), leading to incomplete and unreliable datasets [8–10]. Additionally, inadequate local capacity, reliance on external funding, and poorly developed health information systems are also bringing challenges to national health data availability in LMICs [11]. Due to these limitations in data availability, global health reporting relies heavily on statistical estimates including imputations to make these data suitable for summarizing health trends and enabling cross-country comparisons [6].

There are several strategies available to handle missing values in health-related data. Methods such as Complete Case Analysis (CCA) and Single Imputation (SI) methods which include mean or median imputation, exclusion and interpolation, and regression-based single imputation are relatively easier to implement [12, 13]. However, these

do not account for uncertainty in imputed data [14]. Multiple imputation techniques such as Iterative Robust Model-based Imputation (IRMI), AMELIA algorithm based on the expectation–maximization with a bootstrapping (EMB), sequential imputation (IMPSEQ), and Multivariate Imputation by Chained Equation (MICE) are also popular methods used for addressing the missing values particular to the situations in the health data analysis [15]. Recently machine learning imputation techniques such as missForest (A random forest-based method) and k Nearest Neighbor (k-NN), and seasonal decomposition methods to handle missing values in time series data are also in use for handling missing values in health research [16]. Multiple imputation has several advantages over other methods, as it duly accounts for uncertainty in imputed data and has flexibility with regard to underlying assumptions [14]. Multiple Imputation methods can handle missing values in data which are based on the assumption that data are Missing At Random (MAR) [17].

The international databases collect national health data from countries through household surveys, civil registration systems, health facility data, topic-specific surveys, and statistical reports. In this context, Missing at Random (MAR) and Missing Not at Random (MNAR) are likely to occur in the national health data. MNAR occurs when missingness is related to the unobserved value itself, such as intermittent scheduling of national surveys, data-sharing policy restrictions and irrelevance in the context of individual countries [18] prevent in data collection or reporting. MCAR is less common in health reporting but may occur when data are randomly unreported without any underlying pattern. In national health datasets, MCAR may be seen if an observatory fails to update data for certain countries or indicators despite the availability of the data, as this missingness is unrelated to specific data characteristics. Possibly, we can assume that many of these missing values are related to other observed values. Therefore, missing data are commonly assumed to be missing at random rather than the other two missing mechanisms namely Missing Completely At Random (MCAR) and Missing not at random (MNAR) [19]. Missing data mechanisms indeed exist on a continuum between MAR and MNAR, and rather than aiming for pure categories, it's more practical to evaluate whether or not any assumption violations impact results meaningfully [20]. Further, the inclusion of variables that likely predict the missing information (Auxiliary variables) in the imputation model increases the likelihood of the MAR assumption being met [21]. The MICE also known as Fully Conditional Specification (FCS) is a widely used reliable multiple imputation method in handling missing values [22]. Previous studies have shown

that the multiple imputation methods including the MICE method provide unbiased estimates for a higher proportion of missing data even up to 90% missingness [16, 23, 24]. However, there is scant literature available on the robustness of the MICE method for handling health data like mortality indicators especially when there are missing proportions in varying amounts (as high as 90% missing).

The present study aimed to provide a preliminary guideline for the choice of the extent of missing proportion to impute with the MICE procedure. The study further examines the robustness of the MICE method in imputing the longitudinal health datasets with missing rates ranging from 10 to 90%. To accomplish this, the study used complete data on mortality-related health indicators including Adolescent Mortality Rate (AMR), Under-five Mortality Rate (UMR), Infant Mortality Rate (IMR), Neonatal Mortality Rate (NMR) and Stillbirth Rate (SBR). We chose these indicators as a representative of the national health data due to the completeness of the data as they are the prime measures of health outcomes and availability of the complete data. We generated random missing values of varying proportions (10–90%) in the complete data and then imputed them using the MICE method to assess its performance in handling missing values of different proportions. The present study is a substantial groundwork and a crucial step towards overcoming the challenges of handling missing data information faced by the research community while utilizing longitudinal databases of international organizations like WHO, UNICEF and alike. This can lead to better decision-making and resource allocation in health policy and planning.

Methodology

Data and data source

The study utilized complete data on mortality-related health indicators namely Adolescent Mortality Rate (AMR), Under-five Mortality Rate (UMR), Infant Mortality Rate (IMR), Neonatal Mortality Rate (NMR) and Stillbirth Rate (SBR). These indicators were prioritized and standardized within the category of ‘Mortality by age’, as outlined in the Global Reference List (GRL) of core health indicators by the World Health Organization (WHO) [25]. Data on these health indicators for selected 100 countries between the period 2015–2019 were extracted from the Global Health Observatory (GHO) database (<https://www.who.int/data/gho>), an interface by the WHO showing health-related metrics which is freely available in the public domain [1]. The countries were selected purposively from a list of 189 countries in the United Nations Development Program (UNDP) Human Development Index (HDI) Report 2019 [26]. Specifically,

25 countries were chosen from each of the four categories based on their HDI values: very high, high, medium, and low HDI levels. Furthermore, the selection of indicators was contingent upon the availability of the complete data for the selected 100 countries over the study period (2015–2019). Consequently, the indicator of Adult Mortality Rate listed under the ‘Mortality by age’ category in the GRL with missing values for the selected 100 countries were excluded. The operational definitions for mortality related health indicators utilized in the study can be found additional file [see Additional file 1]. The health indicator data were organized in a long format to align with the longitudinal nature of the data, wherein each country would have multiple records based on the ‘time’ variable [27].

Study procedure

The present study aimed to assess the robustness of the MICE method for handling longitudinal health datasets with differing missing rates ranging from 10 to 90%. The complete data organized in a long format were utilized for the amputation procedure. We followed a stepwise univariate amputation procedure [28] to generate missing values in the complete dataset. In which, the missing values were randomly generated using RAND function in Microsoft Excel 2019 one variable at a time and the procedure was repeated for all the mortality-related health indicator variables. Since missingness in one variable is independent of the missing value itself and observed values, we classify this missing mechanism as Missing Completely At Random (MCAR). For instance, to get a 10% incomplete dataset from the complete dataset, we generated 10% missing values randomly in each variable (AMR, UMR, IMR, NMR and SBR in the present study) one at a time. Since we randomly generated missing values one variable at a time, it is likely to follow mixed missing patterns where missingness occurs in combination of intermittent, and monotone patterns [29]. The study generated nine incomplete datasets of varying proportions of missing values ranging between 10 and 90% (that is, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% and 90% missing data) from the complete dataset. The amputation process was implemented using Microsoft Excel, 2019. The variable details of complete and generated nine incomplete datasets of varying missing rates have been summarized in Additional file 2 [see Additional file 2].

Thereafter, we included auxiliary variables (those variables which are not in the substantive interest but are suspected to contain the information about the missing information in the variables of interest (here, mortality-related health indicators) [20]) to the imputation model of MICE. We carefully selected auxiliary variables that have complete data or with only minimal missingness.

Auxiliary variables are included in the imputation model to improve prediction quality by capturing relationships that help explain missingness [20, 30, 31]. When these auxiliary variables have missing values, their ability to predict values of primary variables diminishes, leading to less accurate and biased imputations due to insufficient information. The details of auxiliary variables used for the study are summarized in an additional file [see Additional file 3]. Subsequently, the nine incomplete datasets were imputed using the MICE method to generate nine complete imputed datasets using the ‘mice’ package in R version 4.2.0 [32]. The study utilized the Predictive Mean Matching (PMM) tool of MICE to generate imputed values that follow the distribution pattern as that of the available information [33]. For the austerity of the imputation method, we used the standard MICE to impute missing values. Since we included a cross-sectional element variable (‘country’ in the present study), time element variable (‘year’ in the present study) along with auxiliary variables and used ‘Predictive Mean Matching (PMM)’ tool to predict the missing values in the imputation model, the standard MICE expected to address the missingness in the longitudinal structure of the data. Previous literature has recognized this flexibility of standard MICE as an effective imputation method for longitudinal data although it is not particularly a longitudinal imputation method [34, 35]. Another important parameter in the MICE imputation model is the number of imputations (m), has been kept as five ($m=5$). For instance, after applying the MICE method for a dataset with 10% missingness, it would generate a total of 5 different imputed complete datasets. The choice of a fixed minimum number of imputations ($m=5$) was made to align with the default setting in the ‘mice’ package [32]. Additionally, five imputations are often cited as sufficient for moderate levels of missingness [20, 36, 37]. However, we compared evaluation metrics of the 90% imputed data using five imputations ($m=5$) versus twenty imputations ($m=20$) and observed a minimal deviation [see Additional file 6]. Further statistical testing on these multiple imputed dataset would give 5 different estimations to account for the uncertainty, therefore taking an average of these different estimates would provide an unbiased estimate for the missing value [22]. Since the present study was not intended to conduct statistical testing or uncertainty analysis on the imputed datasets to derive statistical inferences, we used the averaging method across multiple imputed datasets to obtain a single pooled dataset. This approach was chosen as a practical solution, considering the computational intensity required to compare the robustness of complete imputed datasets across different levels of missingness. Eventually, the study yielded a total of 10 datasets: one complete dataset and nine imputed

datasets with different missing criteria ranging from 10 to 90%. The study then assessed the robustness of the MICE method by comparing the imputed datasets with the dataset comprising true known values. The details of the study procedure are delineated in Fig. 1.

Robustness assessments of the MICE method in imputing missing values

We assessed the robustness of the MICE method in imputing varying proportions of missing values using two approaches.

Approach I: Repeated Measures Analysis of Variance (RM ANOVA)

We employed RM-ANOVA to assess whether there is any significant difference between the health indicator data of complete and imputed datasets [38]. The variance could be biased due to imputation, potentially violating RM-ANOVA assumptions; therefore, we used the Greenhouse–Geisser correction when reporting RM-ANOVA results. Ideally, we expect the acceptance of the null hypothesis that there is no significant difference between complete and imputed datasets. However, in the event of a significant difference, we reported Bonferroni adjusted multiple comparison findings to specifically determine which imputed dataset, out of the nine datasets with varying proportions of missing data (ranging from 10 to 90%), exhibited a significant difference from the actual dataset with 0% missingness [39].

Approach II: Evaluation metrics

Further, the study estimated evaluation metrics such as Root Mean Square Error (RMSE), Mean Absolute Deviation (MAD), Bias and Proportionate Variance (PV) to examine the robustness of the MICE procedure by assessing and comparing the deviation of imputed data of different missing rates and complete data of mortality related health indicators. We estimated the following evaluation metrics:

Root mean square error (RMSE)

The study calculated Root Mean Squared Error (RMSE) for all imputed datasets with different missing rates ranging from 10 to 90% using the following equation [16].

$$RMSE = \sqrt{\frac{\sum (x_{imputed} - x_{complete})^2}{n}}$$

where, $x_{imputed}$ is the imputed value, $x_{complete}$ is the actual value corresponding to the imputed value and n is the number of cases imputed. The RMSE estimate the difference between the true/complete value and the imputed value, and it is considered as a measure of the bias. The smaller RMSE indicates less bias and more consistency after imputation. Hence, the RMSE value close to zero is

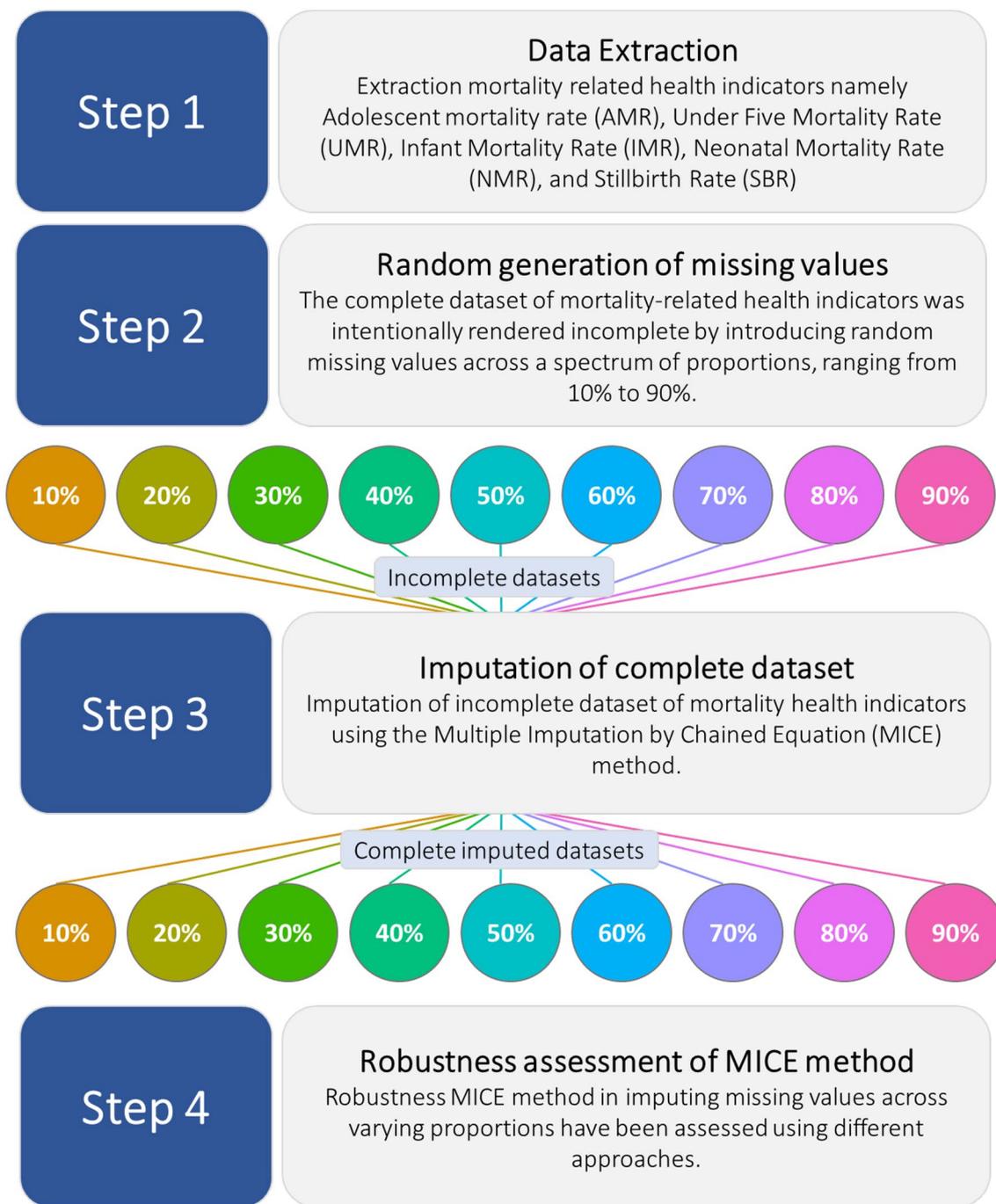


Fig. 1 Study flow chart

desirable. Additionally, we estimated the relative RMSE value by dividing the RMSE of the imputed data by the standard deviation of the complete/true dataset. The details are given in the Additional file 5 [see Additional file 5].

Mean absolute deviation (MAD)

It is an evaluation metric which estimates the average of the absolute difference between the imputed value and complete value of mortality-related indicators. It is one of the measures to know dispersion in the imputed data

from the complete data given by the following formula [40].

$$MAD = \frac{\sum |x_{imputed} - x_{complete}|}{n}$$

The smaller MAD indicates less bias and more consistency after imputation, hence, a value closer to zero is the desirable value.

Bias

It is an evaluation metric which estimates the mean deviation between the imputed value and the complete value of mortality-related indicators. The bias is estimated using the following formula [16].

$$BIAS = \frac{\sum (x_{imputed} - x_{complete})}{n}$$

A value of zero indicates no bias. A positive bias indicates the overestimation of the complete data and a negative value indicates the underestimation of the complete data by the imputed values.

Proportionate variance (PV)

A proportionate variance is the ratio of the variance of the imputed values to the variance of the corresponding complete data. The PV values help to assess the extent of the imputed value and capture the variance of the complete data. The PV is calculated using the following equation [40]:

$$PV = \frac{\text{variance}(x_{imputed})}{\text{variance}(x_{complete})}$$

The PV value of 1 implies that the variance of the imputed and complete data is equal. The PV value closer to 1 is desirable.

The estimated evaluation metrics were categorized as 'high performance', 'medium performance' and 'Low/cautionary performance' for each health indicator. The categorization was based on the observed range for each evaluation metric divided into equal parts, aiming to create distinct categories based on the observed spread in the data. The details of the categorization of evaluation metrics for each health indicator have been described in the additional file [see Additional file 4]. We examined the robustness of the imputed datasets of all the indicators from the complete data by assessing the performance in each evaluation metric.

Approach III: Visual inspection of the box plots

It is crucial to scrutinize other key statistical estimates such as interquartile range, maximum and minimum values, and the presence of outliers across the complete and imputed datasets. This comprehensive evaluation allows

us to assess the imputation's efficacy in capturing not only the mean but also the variability present in the original data. To facilitate this analysis, we utilized box plots to visually compare the distributions of health indicators between the complete and imputed datasets [16].

Result

Table 1 details the test of RM-ANOVA of various indicators of complete and imputed datasets. The mean and standard deviation of each mortality-related health indicators of selected 100 countries over the period 2015–2019 of both complete and imputed datasets of different missing criteria (10–90%) were reported. Additionally, we have reported the mean difference of indicators of imputed datasets from the complete dataset. The RM-ANOVA evaluated the null hypothesis that there is no significant difference between mortality-related health indicators of the complete and imputed datasets. Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated, for all the indicators under study including AMR ($\chi^2=10,160.808$, $p<0.001$), UMR ($\chi^2=9940.261$, $p<0.001$), IMR ($\chi^2=9723.733$, $p<0.001$), NMR ($\chi^2=10,364.271$, $p<0.001$), and SBR ($\chi^2=12,384.048$, $p<0.001$). Therefore, a Greenhouse–Geisser correction was used to report RM-ANOVA. The RM-ANOVA, with Greenhouse–Geisser correction, revealed statistically significant differences in the mean values of AMR ($F(2.059, 1027.546)=4.864$, $p=0.007$), UMR ($F(2.531, 1263.192)=19.551$, $p<0.001$), IMR ($F(3.141, 1567.604)=6.818$, $p<0.001$), NMR ($F(2.248, 1121.978)=3.577$, $p=0.024$), and SBR ($F(2.485, 1239.783)=3.459$, $p=0.023$) between the complete and imputed datasets. Since there is a significant difference between imputed and complete datasets we conducted Bonferroni adjusted multiple comparison analysis [39]. Table 1 provides a detailed pairwise comparison of each indicator between the complete and imputed datasets. A statistically significant difference was experienced for 80%, 90%, 60% and 70% imputed data of AMR (mean difference of 6.419 deaths per 100,000 adolescent population), UMR (mean difference of 7.424 deaths per 1000 live births), IMR (mean difference of 2.095 deaths per 1000 live births) and SBR (mean difference of 0.745 stillbirths per 1000 total births) indicators in comparison to the dataset without imputation. However, NMR did not show a statistically significant difference between the non-imputed and imputed datasets with varying missing criteria.

The evaluation metrics such as Root Mean Square Error (RMSE), Mean Absolute Deviation (MAD), Bias and Proportionate Variance (PV) were used to assess the extent of deviation of the imputed datasets (with missing criteria ranging from 10 to 90%) from the complete

Table 1 Details of the test of RM-ANOVA and post-hoc analysis of mortality-related health indicators of 10–90% imputed and complete datasets

Imputed datasets with missing criteria	Mean ± SD	Range	Mean difference from complete data	95% CI		p value
				Lower limit	Upper limit	
Adolescent mortality rate (AMR)						
Mauchly's test: Chi-square (χ^2): 10,160.808, <i>p</i> value: < 0.001			RM-ANOVA: F (<i>df</i> _{group} , <i>df</i> _{error}): ^b F (2.059, 1027.546) = 4.864, <i>p</i> value = 0.007			
0%	89.02 ± 72.12	10.61–304.13				
10%	89.05 ± 72.01	10.61–304.13	−0.025	−0.22	0.17	1
20%	89.08 ± 72.15	10.61–297.19	−0.056	−0.324	0.212	1
30%	89.95 ± 71.84	10.61–297.19	−0.922	−2.002	0.157	0.238
40%	88.48 ± 71.86	10.61–297.19	0.543	−0.189	1.275	0.69
50%	86.97 ± 70.43	12.12–297.19	2.056	−0.07	4.183	0.073
60%	88.14 ± 71.71	10.61–297.19	0.886	−0.654	2.426	1
70%	88.14 ± 63.9	11.64–286.47	0.887	−4.011	5.785	1
80%	95.17 ± 59.7	12.12–291.87	−6.149	−12.139	−0.158	0.037 ^a
90%	84.68 ± 20.69	12.51–286.47	4.341	−5.791	14.473	1
Under-five mortality rate (UMR)						
Mauchly's test: Chi-square (χ^2): 9940.261, <i>p</i> value: < 0.001			RM-ANOVA: F (<i>df</i> _{group} , <i>df</i> _{error}): ^b F (2.531, 1263.192) = 19.551, <i>p</i> < 0.007			
0%	35.29 ± 32.29	2–140.19				
10%	35.31 ± 32.29	2–140.19	−0.015	−0.085	0.055	1
20%	35.35 ± 32.38	2–140.19	−0.059	−0.211	0.093	1
30%	35.31 ± 32.41	2–140.19	−0.018	−0.215	0.179	1
40%	35.39 ± 31.96	2–124.92	−0.101	−0.506	0.303	1
50%	34.68 ± 31.75	2–128.68	0.614	−0.206	1.434	0.646
60%	34.06 ± 29.89	2.07–124.92	1.229	−0.505	2.963	0.923
70%	34.31 ± 29.23	2–140.19	0.986	−1.14	3.111	1
80%	33.3 ± 26.03	2.21–124.92	1.994	−0.899	4.887	1
90%	27.87 ± 16.35	2.44–106.62	7.424	3.406	11.442	< 0.001 ^a
Infant mortality rate (IMR)						
Mauchly's test: Chi-square (χ^2): 9723.733, <i>p</i> value: < 0.001			RM-ANOVA: F (<i>df</i> _{group} , <i>df</i> _{error}): ^b F (3.141, 1567.604) = 6.818, <i>p</i> < 0.001			
0%	26.21 ± 21.93	1.59–95.12				
10%	26.21 ± 21.92	1.59–95.12	0	−0.056	0.055	1
20%	26.22 ± 21.89	1.59–95.12	−0.005	−0.103	0.094	1
30%	26.19 ± 21.91	1.59–89.73	0.016	−0.072	0.105	1
40%	26.39 ± 21.77	1.65–89.73	−0.175	−0.506	0.155	1
50%	26.32 ± 21.69	1.59–89.73	−0.108	−0.648	0.432	1
60%	24.12 ± 19.37	1.7–87.25	2.095	0.611	3.579	< 0.001 ^a
70%	25.97 ± 19.5	1.82–85.63	0.245	−1.216	1.706	1
80%	24.55 ± 17.21	1.76–87.93	1.665	−0.331	3.66	0.29
90%	27.23 ± 14.09	2.06–87.25	−1.023	−3.497	1.451	1
Neonatal mortality rate (NMR)						
Mauchly's test: Chi-square (χ^2): 10,364.271, <i>p</i> value: < 0.001			RM-ANOVA: F (<i>df</i> _{group} , <i>df</i> _{error}): ^b F (2.248, 1121.978) = 3.577, <i>p</i> = 0.024			
0%	15.56 ± 11.7	0.85–45.25				
10%	15.55 ± 11.71	0.85–45.25	0.01	−0.009	0.029	1
20%	15.56 ± 11.72	0.85–45.25	0	−0.032	0.031	1

Table 1 (continued)

Imputed datasets with missing criteria	Mean ± SD	Range	Mean difference from complete data	95% CI		p value
				Lower limit	Upper limit	
30%	15.55 ± 11.71	0.86–45.25	0.008	−0.058	0.074	1
40%	15.27 ± 11.47	0.85–45.25	0.287	−0.065	0.639	0.344
50%	15.3 ± 11.26	0.87–45.09	0.264	−0.201	0.728	1
60%	15.09 ± 11.37	0.85–45.25	0.468	−0.001	0.937	0.051
70%	15.94 ± 11.07	0.87–45.09	−0.382	−0.995	0.23	1
80%	14.57 ± 8.93	0.99–40.46	0.994	−0.124	2.112	0.167
90%	15.05 ± 4.24	0.85–45.07	0.513	−1.117	2.142	1
Stillbirth rate (SBR)						
Mauchly's test:			RM-ANOVA:			
Chi-square (χ^2): 12,384.048, p value: < 0.001			F (df_{group}, df_{error}): ^b F (2.485, 1239.783) = 3.459, p = 0.023			
0%	12.7 ± 8.68	1.52–34.5				
10%	12.7 ± 8.68	1.52–34.5	0.002	−0.015	0.019	1
20%	12.69 ± 8.67	1.56–34.5	0.003	−0.016	0.023	1
30%	12.68 ± 8.66	1.52–34.5	0.014	−0.013	0.042	1
40%	12.69 ± 8.68	1.52–34.5	0.007	−0.036	0.051	1
50%	12.83 ± 8.63	1.71–34.5	−0.131	−0.294	0.031	0.376
60%	12.48 ± 8.22	1.52–32.7	0.214	−0.209	0.637	1
70%	11.95 ± 7.57	1.56–31.47	0.745	0.106	1.384	0.007 ^a
80%	12.78 ± 7.1	1.7–34.5	−0.08	−0.828	0.668	1
90%	12.51 ± 5.24	1.79–31.47	0.132	−0.828	0.668	1

^a Significance level at 0.05

^b Greenhouse- Geisser corrected F-statistic

dataset (non-imputed) of mortality related health indicators has been presented in Table 2 and Fig. 2. The highest RMSE values (RMSE value of AMR- 72.9; UMR- 29.9; IMR- 17.8; NMR- 11.7; and SBR- 7.4) were observed for 90% imputed data for all the health indicators (Table 2 and Fig. 2A). Likewise, at lower missing data levels (10–30%), relative RMSE values remain low, but as missing data exceeds 50%, the Relative RMSE increases, indicating greater deviation of imputed values from the original data's variability (see Additional file 5). A similar trend was observed for MAD that, the highest values (MAD values of AMR- 58.7; UMR- 20.2; IMR- 13.3; NMR- 10.1, and SBR- 5.2) were observed for 90% imputed data (Table 2 and Fig. 2B). Moreover, the RMSE and MAD values for SBR were below 10 units across all imputed datasets, whereas for NMR, these values remained below 10 units up to the 80% imputed dataset. The MAD values were below 10 units for AMR, UMR and IMR up to 60%, 70% and 80% imputed datasets respectively.

For the UMR, IMR, NMR, and SBR indicators, the highest bias values were 8.3 (90% imputed data), 3.5 (60% imputed data), 1.2 (80% imputed data), and 1.1 (70% imputed data), respectively, all of which were observed to underestimate the corresponding complete values.

In contrast, the highest bias value for the AMR indicator was 7.7 with 80% imputed data, which overestimated the complete values. The majority of mortality-related indicators of various imputed datasets exhibited negative bias values, suggesting an underestimation of the complete data by the imputed data (Table 2 and Fig. 2C). As the proportion of missing data increased, the deviation in PV values from the ideal score of 1 also showed a rising trend (Table 2 and Fig. 2D). The highest deviation from the ideal value was observed with 90% imputed data for all indicators. Overall, the evaluation metrics indicated a decrease in robustness as the proportion of missing data increased.

Figure 3 depicts the performance level of the MICE method based on each evaluation metric for each health indicator of nine imputed datasets. The evaluation metrics predominantly exhibited 'high performance' up to the dataset with a 50% missing proportion for various health indicators. Further, for the datasets with the imputation of 60–70% missing proportion, most evaluation metrics of different indicators showed a mix of 'high' and 'medium' performance levels; although a few evaluation metrics showed 'low performance' levels as well. However, with missing proportions exceeding 70%, the

Table 2 Evaluation metrics of imputed data with varying missing proportions of various health indicators in comparison with complete data

Imputed datasets with missing criteria	Adolescent mortality rate	Under-five mortality rate	Infant mortality rate	Neonatal mortality rate	Stillbirth rate
Root Mean Square Error—RMSE (<i>RMSE value close to zero is desirable</i>)					
10%	4.205 ^a	1.509 ^a	1.201 ^a	0.406 ^a	0.366 ^a
20%	4.082 ^a	2.323 ^a	1.500 ^a	0.485 ^a	0.300 ^a
30%	13.531 ^a	2.453 ^a	1.099 ^a	0.819 ^a	0.346 ^a
40%	7.930 ^a	4.361 ^a	3.574 ^a	3.817 ^a	0.466 ^a
50%	20.691 ^a	7.944 ^a	5.204 ^a	4.489 ^b	1.577 ^a
60%	13.587 ^a	15.328 ^b	13.324 ^c	4.165 ^b	3.728 ^b
70%	39.886 ^b	17.341 ^b	11.897 ^c	5.007 ^b	5.276 ^c
80%	46.129 ^b	22.143 ^c	15.310 ^c	8.586 ^c	5.697 ^c
90%	72.885 ^c	29.888 ^c	17.796 ^c	11.712 ^c	7.385 ^c
Mean Absolute Deviation—MAD (<i>MAD value close to zero is desirable</i>)					
10%	2.961 ^a	1.049 ^a	0.785 ^a	0.342 ^a	0.263 ^a
20%	3.006 ^a	1.538 ^a	1.053 ^a	0.395 ^a	0.228 ^a
30%	5.768 ^a	1.743 ^a	0.770 ^a	0.507 ^a	0.229 ^a
40%	4.311 ^a	1.990 ^a	1.326 ^a	1.206 ^a	0.298 ^a
50%	7.231 ^a	2.653 ^a	2.041 ^a	1.559 ^a	0.499 ^a
60%	6.755 ^a	6.055 ^a	4.980 ^b	1.638 ^a	1.581 ^a
70%	17.243 ^a	9.157 ^b	5.511 ^b	2.285 ^a	2.526 ^b
80%	29.143 ^b	12.672 ^b	8.752 ^b	5.033 ^b	3.291 ^b
90%	58.671 ^c	20.224 ^c	13.267 ^c	10.131 ^c	5.217 ^c
BIAS (<i>BIAS value of zero indicates no bias, a positive value indicates the overestimation and a negative value indicates an underestimation of complete data</i>)					
10%	0.248 ^a	0.149 ^a	0.003 ^a	-0.097 ^a	-0.016 ^a
20%	0.279 ^a	0.295 ^a	0.023 ^a	0.002 ^a	-0.016 ^a
30%	3.074 ^b	0.059 ^a	-0.054 ^a	-0.026 ^a	-0.048 ^a
40%	-1.358 ^a	0.254 ^a	0.439 ^a	-0.719 ^b	-0.018 ^a
50%	-4.113 ^b	-1.228 ^a	0.216 ^a	-0.527 ^b	0.262 ^a
60%	-1.476 ^a	-2.048 ^a	-3.492 ^c	-0.780 ^b	-0.357 ^b
70%	-1.268 ^a	-1.408 ^a	-0.350 ^a	0.546 ^b	-1.065 ^c
80%	7.686 ^c	-2.493 ^a	-2.081 ^b	-1.242 ^c	0.100 ^a
90%	-4.823 ^b	-8.249 ^c	1.137 ^a	-0.570 ^b	-0.203 ^a
Proportionate Variance—PV (<i>The PV value closer to 1 is desirable</i>)					
10%	0.968 ^a	1.002 ^a	0.992 ^a	1.005 ^a	1.010 ^a
20%	1.005 ^a	1.056 ^a	0.984 ^a	1.014 ^a	0.995 ^a
30%	0.984 ^a	1.023 ^a	0.994 ^a	1.003 ^a	0.989 ^a
40%	0.977 ^a	0.953 ^a	0.965 ^a	0.901 ^a	1.004 ^a
50%	0.886 ^a	0.928 ^a	0.955 ^a	0.864 ^a	0.978 ^a
60%	0.981 ^a	0.774 ^a	0.690 ^b	0.912 ^a	0.838 ^a
70%	0.698 ^a	0.743 ^a	0.710 ^b	0.849 ^a	0.664 ^b
80%	0.613 ^b	0.565 ^b	0.514 ^c	0.495 ^c	0.590 ^b
90%	0.002 ^c	0.208 ^c	0.338 ^c	0.018 ^c	0.299 ^c

^a High performance; ^b medium performance; ^c low/cautionary performance

The categorization details of the performance levels of evaluation metrics are presented in an additional file [see Additional file 4]

majority of indicators demonstrated a ‘low’ performance level in terms of most evaluation metrics.

Figure 4, delineates the box plot for each mortality-related indicator and contains a plot for both complete

and imputed data. The visual inspection of the box plots suggests that till a missing proportion of 50%, the imputed data is slightly changing from the complete data. Whereas, the imputed data for the missing proportion of

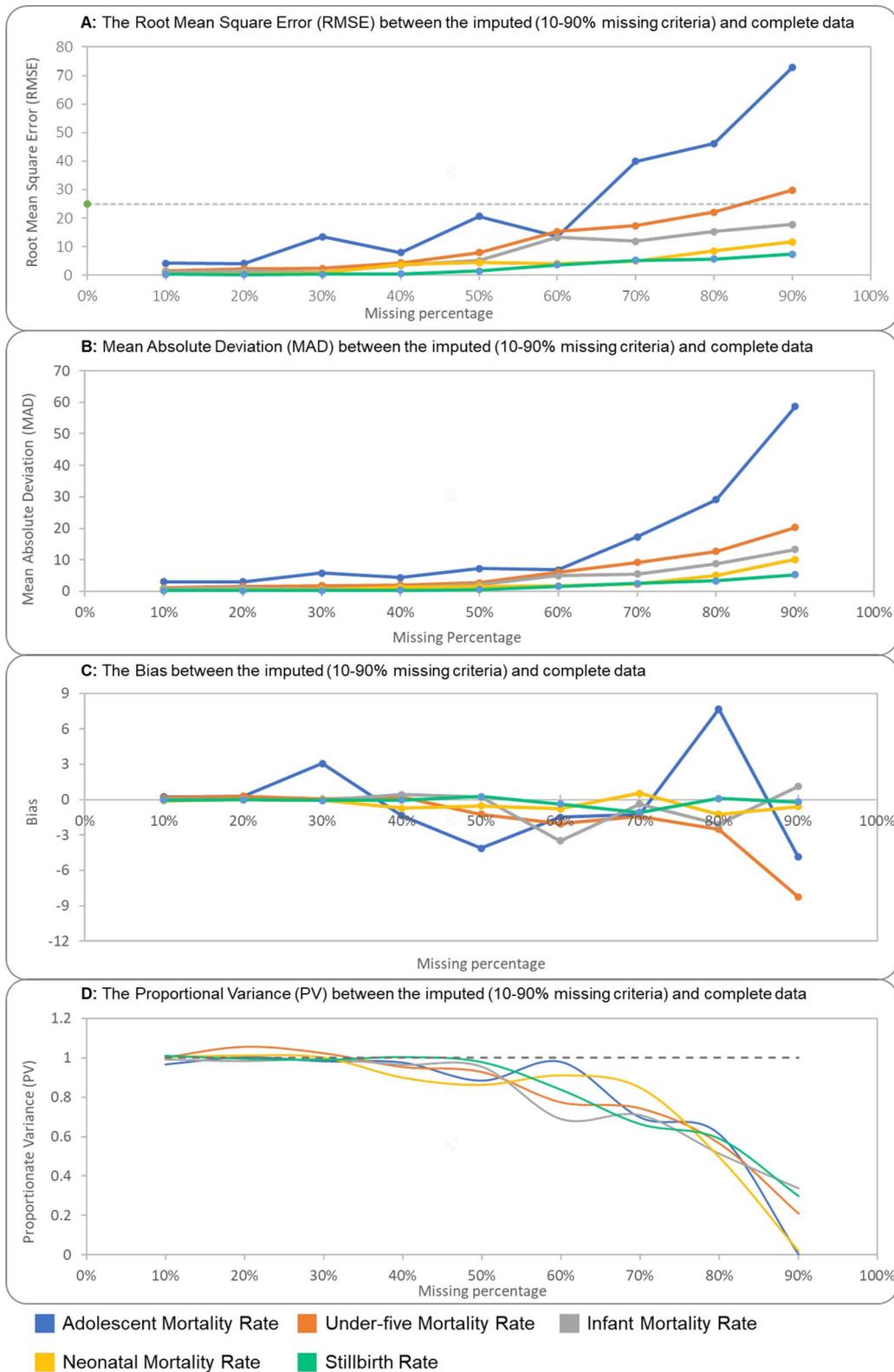


Fig. 2 Trends in evaluation metrics of mortality-related health indicators of imputed datasets differing in missing criteria (10–90%)

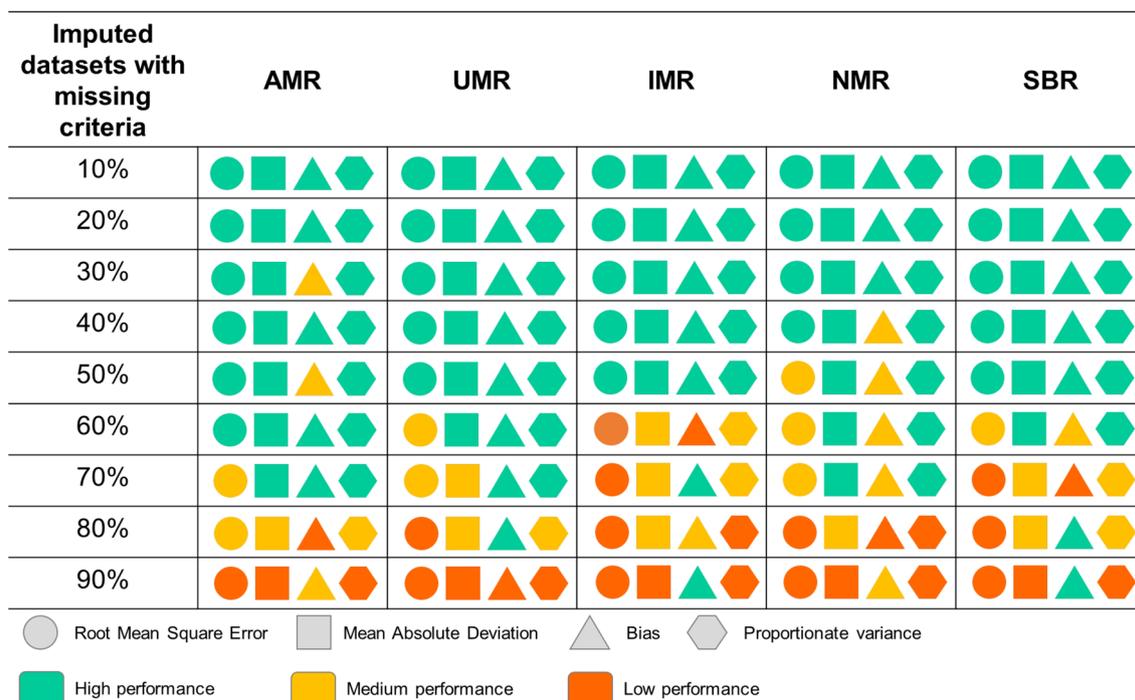


Fig. 3 Performance levels of different MICE imputed datasets with regard to evaluation metrics

60% and 70% shows a moderate change from the complete data and that of imputed data with more than 70% missing rates shows substantial changes from the complete data.

Discussion

Despite MICE being recognized as a robust imputation method, there remains a lack of guidelines regarding the acceptable proportion of missing data that can be effectively imputed using this method without significantly compromising data accuracy and reliability. Graham (2009) suggested that multiple imputation, including MICE, can effectively handle up to 50% of missing data, even in datasets with small sample sizes [20]. Conversely, Kim & Kim (2020) observed acceptable performance by the MICE method for datasets with missing proportions of up to 60% in their simulation study [41]. Similarly, Kambach et al. (2020) observed in their simulation study that MICE can handle missing proportions up to 90% with relatively less bias and better accuracy compared to other imputation methods [42]. Blazek et al. (2021) also noted that multiple imputation can provide unbiased estimates of missing values even when the missing proportions are very high with the inclusion of appropriate auxiliary variables, which are associated with missing data, in the imputation model [23]. Taking cue from literature, it can be derived that MICE can handle missing proportions as high as 90%, however; to what extent the

performance levels of different MICE imputed datasets can be labelled as ‘high’, ‘medium’ and ‘cautionary’ based on evaluation metrics?

Against this backdrop, the study aimed to assess the robustness of the MICE method in imputing health indicators with varying missing proportions by using four different statistical approaches. Specifically, it sought to determine the extent to which MICE could effectively handle missing values in longitudinal health data. Using the first approach, the study compared the means of imputed datasets with the complete data of health indicators using the RM-ANOVA procedure. Previous literature has also used one-way ANOVA and subsequent post-hoc tests to compare the performance of different methods of imputation [40]. Upon conducting the pairwise comparison between each imputed data and corresponding complete data, no statistically significant difference was observed for any health indicator when imputation was performed up to 50% missing proportions. This finding suggests that imputation up to the threshold of 50% missing proportion of similar data can yield comparable results to the complete data.

The evaluation metrics (second approach) used in this study—RMSE, MAD, Bias, and PV—were essential for assessing the robustness of imputed datasets across varying missing data scenarios, from 10 to 90%. Generally, the findings revealed a consistent pattern: evaluation metrics tended to increase as the proportion of missing

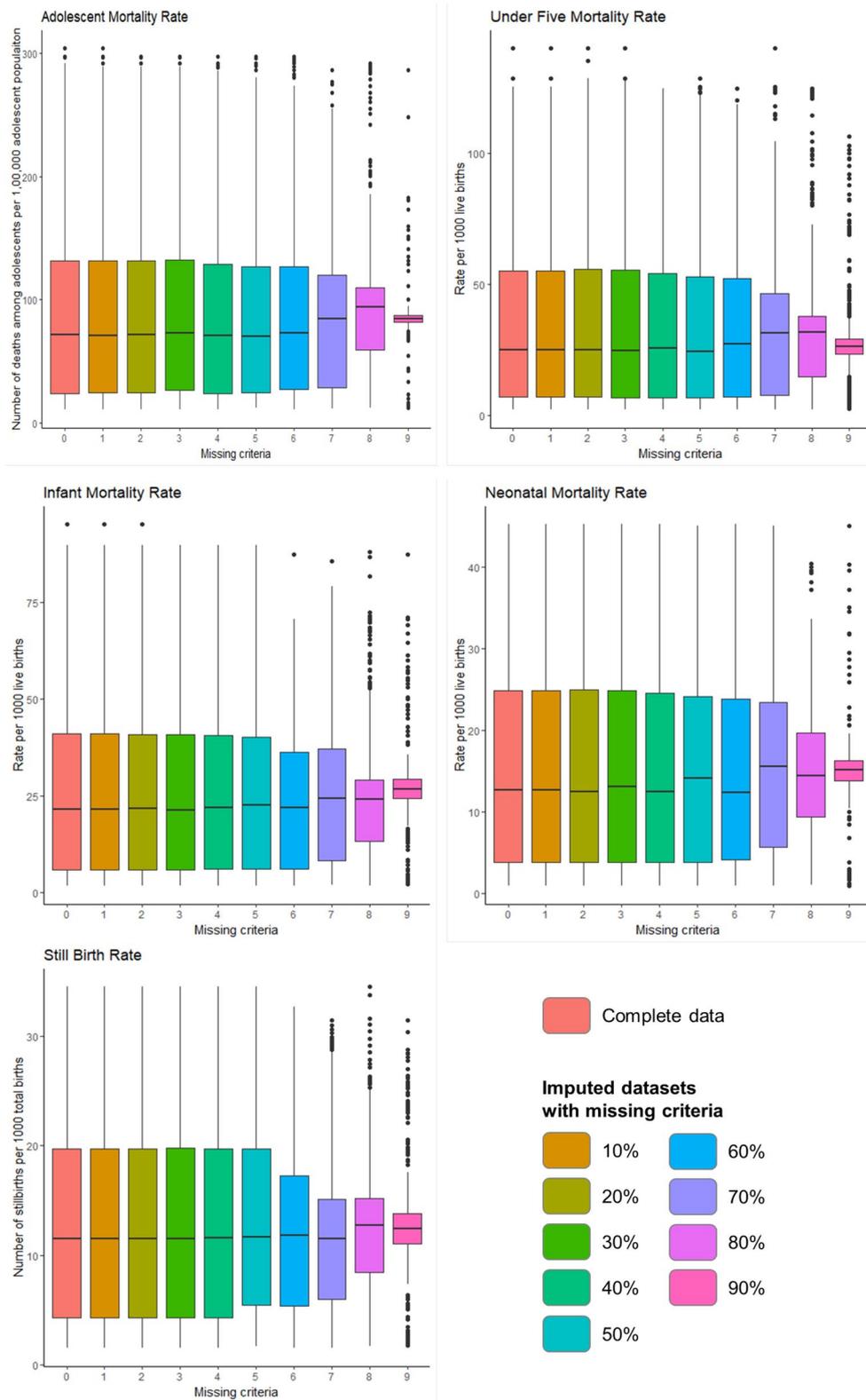


Fig. 4 Boxplot of the complete and imputed data (10–90% missing proportions) of mortality-related health indicators

data imputed grew. Additionally, we estimated the relative RMSE to have a sense of imputation error relative to the natural variability. It also shows that, as the missing proportion increases, the relative RMSE values rise, particularly beyond 50%, suggesting higher imputation error relative to natural variability at these levels. This pattern aligns with observations from past literature, which have shown similar trends in the plots between evaluation metrics such as crude RMSE and crude BIAS and the proportions of missing data [16]. In the present study, the bias values for most mortality-related indicators were negative, suggesting an underestimation of the complete data by the imputed data. In the Feng et al. (2021) study, the bias for the best-performing imputation method ranged between -2.5 to 1 when handling data with a missing proportion ranging between 5 and 30% [16]. In line with this, the bias value ranged from -2 to 1 for all the indicators except the indicator of AMR up to 50% missing proportion in the present study (Table 2). It became evident from the present study that PV values deviated further from the ideal value of 1 as the proportion of imputed missing data increased. A notable deviation in PV values was particularly observed beyond the 70% missing data criterion, indicating a significant underestimation of the variance compared to the complete dataset as the proportion of missing values increased (Table 2). Similarly, the study identified significant shrinkage in the standard deviation particularly beyond 70% missingness, mirroring the trends observed in PV values. It is crucial to interpret pairwise comparison findings between the complete and imputed data alongside the visual inspection of evaluation metrics and box plots (Figs. 3 and 4). For instance, while the 90% imputed AMR indicator did not exhibit a significant difference from the complete data upon initial inspection, however; examination of the evaluation metrics plot (Figs. 2 and 3) and box plot (Fig. 4) revealed substantial deviation, particularly after reaching a 70% missing proportion. Therefore, to accurately represent both the mean and variance of the imputed data and the complete data, visual inspection of box plots was also resorted to in the present study. The examination of evaluation metrics (Fig. 3) and box plots (the third approach) (Fig. 4) revealed a decline in robustness as the proportion of missing data increased. In general, up to 50% missing data, deviations of imputed data with complete data were minimal, while beyond this threshold, some indicators showed considerable deviations. Furthermore, as missing proportions surpassed 70% , most indicators exhibited significant disparities between the imputed and complete non-imputed data.

The present study has several strengths to offer. The literary evidence [43–45] found MICE to be robust, providing unbiased estimates with minimal prediction error

compared to other imputation methods. Nonetheless, its effectiveness in imputing data with varying degrees of missingness especially in population health metrics, such as mortality, fertility, health systems, etc. indicators, remains relatively unexplored, the lacunae which have been filled by the present study. In this context, it is the first of its kind, to the best of the authors' knowledge, to assess the robustness of the MICE method in imputing missing data proportions ranging from 10 to 90% and comparing these imputations with the complete dataset. The study has followed a very detailed exhaustive procedure by using four different statistical approaches to provide a preliminary guideline for determining the choice of the extent of imputing data that can be robustly handled by the MICE method. The study has further provided a novel categorization to label the performance levels of different MICE imputed datasets as 'high', 'medium' and 'cautionary' based on multiple reliable evaluation metrics. The present study thus serves as a reference point for overcoming the challenges of handling missing data, especially while utilizing longitudinal databases of the international organization. This can lead to better decision-making and resource allocation in health policy and planning.

The study had several limitations that should be considered. First, we generated an equal proportion of missing values across all indicators, which may not reflect real scenarios. Further research is needed to address this gap. Second, the study focused solely on population metrics related to mortality health indicators. Future studies should explore the robustness of the MICE method in imputing other health indicators. The present study aimed to assess, being a widely used multiple imputation method, how well MICE performs when imputing longitudinal national health datasets with varying missing rates. However, due to the computing intensity, we could not compare the performance of various imputation methods. The study used a simpler stepwise univariate amputation procedure for generating missing values under the condition of 'missing completely at random'. Therefore, readers should be cautioned while drawing inferences from the findings. The present study used a fixed minimum number of imputations ($m=5$) for imputing even for data with a higher proportion of missing (e.g.: 60 – 90%). Five imputations are often cited as sufficient for moderate levels of missingness [20, 36, 37]. Therefore increasing the number of imputations when dealing data with a higher proportion of missingness may reduce the imputation error and increase statistical power [22]. The present study limited its analysis by detecting changes in the overall mean to provide a general view of the dataset's performance after imputation. However, due to the computing intensity, we did not

conduct any per-country analysis. In the present study, we did not conduct a simulation study, which limits the robustness of our evaluation of the MICE imputation method. Since there is no de facto approach available, the categorization of the evaluation metrics performance in the present study has been done by dividing the observed range, therefore readers are advised to interpret these categorizations with caution. Additionally, including more variables in the imputation model could potentially yield better results. Researchers and policymakers must be aware of the potential biases and deviations that may arise from imputation, especially when dealing with a high proportion of missing data (over 70%). Furthermore, the performance of MICE in imputing a higher proportion of missing data is influenced by other factors such as the type and pattern of missing data [46], the number of auxiliary variables, parameters such as number of imputations and iterations and the extent of missingness in the auxiliary variables. Our study aims to offer a foundational understanding, with the current findings serving as a first step toward more extensive investigations. Therefore, the present study provides a preliminary guideline based on the findings from a specific study context. We caution the readers that these guidelines are not definitive; rather, they are intended to offer insight into the thresholds of missing data for MICE based on empirical evidence derived from our analysis.

Future studies should expand upon this research by applying imputation methods to national health datasets under various missingness mechanisms, including MAR, MCAR, and MNAR while addressing diverse missingness patterns such as univariate, monotone, intermittent, and mixed. Simulation studies are recommended to improve the evaluation accuracy of different imputation methods and validate their robustness. Additionally, since this study focuses on aggregated national health data, future research could offer valuable insights through country-specific analyses to better understand imputation performance across different contexts. Exploring the imputation error in cases of high missingness levels is also important, particularly by optimizing the MICE model parameters, such as increasing the number of imputations and iterations. Finally, future research should extend beyond MICE to compare a variety of imputation methods, including multilevel MICE, to determine the most effective approaches for addressing complex missingness patterns in longitudinal national health datasets.

Conclusion

The study offers preliminary guidelines for selecting the appropriate extent of missing data to impute using MICE, based on an evaluation of the MICE method's

performance in handling missing data for mortality health indicators. We found that MICE is effective for imputing missing values up to 50%, showing only marginal deviations from complete datasets. However, for missing proportions between 50 and 70%, moderate alterations occur, and for proportions beyond 70%, significant shrinkage in variance and poor evaluation metric performance compromise data reliability and accuracy. This study underscores the importance of understanding imputation limitations and biases, offering practical guidance for researchers and policymakers. Further research is needed to explore MICE's performance in diverse contexts and improve its robustness for informed decision-making in public health.

Abbreviations

AMR	Adolescent mortality rate
CCA	Complete case analysis
GHO	Global Health Observatory
GRL	Global reference list
HDI	Human Development Index
IMR	Infant mortality rate
k-NN	K-nearest neighbor
MAD	Mean absolute deviation
MAR	Missing at random
MCAR	Missing completely at random
MICE	Multiple imputation by chained equations
MNAR	Missing not at random
NMR	Neonatal mortality rate
PV	Proportionate variance
RMSE	Root mean square error
SBR	Stillbirth rate
SI	Single imputation
UMR	Under-five mortality rate
UNAIDS	United Nations Program on HIV/AIDS
UNDP	United Nations Development Program
UNICEF	United Nations Children's Fund
WHO	World Health Organization

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12963-025-00364-2>.

Additional file 1
Additional file 2
Additional file 3
Additional file 4
Additional file 5
Additional file 6

Acknowledgements

We acknowledge University Grants Commission for providing JRF/SRF fellowship to the first author (KP) for pursuing his Ph.D. degree.

Author contributions

TK and KP conceptualized the study and contributed to the design. KP extracted the data and conducted the statistical analyses. TK and KP contributed to the interpretation of data. KP wrote the first draft of the paper. TK critically revised the first draft. All the authors (TK, KP, MG, KK and SS) reviewed and approved the final manuscript.

Funding

The authors did not receive support from any organization for the publication of submitted work. The first author (KP) is a recipient of University Grants Commission (UGC) – JRF/SRF Fellowship Scheme [No. 1452/(NET-JULY 2018)] for pursuing his Ph.D. degree.

Availability of data material

The datasets used and/or analysed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interest

The authors declare no competing interests.

Author details

¹Department of Community Medicine and School of Public Health, Postgraduate Institute of Medical Education and Research (PGIMER), Chandigarh, India.

²Department of Biostatistics, Postgraduate Institute of Medical Education and Research, Chandigarh, India. ³Department of Obstetrics and Gynaecology, Postgraduate Institute of Medical Education and Research, Chandigarh, India.

Received: 16 July 2024 Accepted: 25 January 2025

Published online: 01 February 2025

References

- WHO. WHO. World Health Organization. <https://www.who.int/data/gho>. Accessed 7 Jan 2021.
- UNICEF. UNICEF Data Warehouse. https://data.unicef.org/dv_index/. Accessed 19 Feb 2021.
- UNAIDS. AIDSinfo. United Nations program on HIV/AIDS. <https://aidsinfo.unaids.org/>. Accessed 7 Jan 2021.
- Hung YW, Hoxha K, Irwin BR, Law MR, Grépin KA. Using routine health information data for research in low- and middle-income countries: a systematic review. *BMC Health Serv Res*. 2020;20:1–15.
- Hoxha K, Hung YW, Irwin BR, Grépin KA. Understanding the challenges associated with the use of data from routine health information systems in low- and middle-income countries: a systematic review. *Heal Inf Manag J*. 2020.
- AbouZahr C. Health Information Systems. In: Raviglione MCB, Tediosi F, Villa S, Casamitjana N, Plasencia A, editors. *Global Health Essentials*. Cham: Springer International Publishing; 2023. p. 303–7. https://doi.org/10.1007/978-3-031-33851-9_46.
- Tuladhar S, Mwamelo K, Manyama C, Obuobi D, Antunes M, Gashaw M, et al. Proceedings from the CIHLMU 2022 Symposium: “Availability of and Access to Quality Data in Health.” *BMC Proc*. 2023;17:21. <https://doi.org/10.1186/s12919-023-00270-1>.
- Hlaing T, Zin T. Organizational factors in determining data quality produced from health management information systems in low-and middle-income countries: a systematic review. *Heal Informatics*. 2020;9:10–5121.
- Burkle FM. Opportunities lost: political interference in the systematic collection of population health data during and after the 2003 War in Iraq. *Disaster Med Public Health Prep*. 2021;15:144–50.
- Silva R, Mizoguchi N. Mortality Data in Service of Conflict-Affected Populations. In: Macfarlane SB, AbouZahr C, editors. *The Palgrave Handbook of Global Health Data Methods for Policy and Practice*. London: Palgrave Macmillan UK; 2019. p. 245–62. https://doi.org/10.1057/978-1-137-54984-6_13.
- Zhao L, Cao B, Borghi E, Chatterji S, Garcia-Saiso S, Rashidian A, et al. Data gaps towards health development goals, 47 low- and middle-income countries. *Bull World Health Organ*. 2022;100:40–9.
- Salgado CM, Azevedo C, Proença H, Vieira SM. Missing Data BT - Secondary Analysis of Electronic Health Records. In: Data MITC, editor. *Cham: Springer International Publishing*; 2016. p. 143–62. https://doi.org/10.1007/978-3-319-43742-2_13.
- McCleary L. Using multiple imputation for analysis of incomplete data in clinical research. *Nurs Res*. 2002;51:339–43.
- Azur M, Stuart E, Frangakis C, Leaf P. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res*. 2011;20:40–9. <https://doi.org/10.1002/mpr>.
- Kabir G, Tesfamariam S, Hemsing J, Sadiq R. Handling incomplete and missing data in water network database using imputation methods. *Sustain Resilient Infrastruct*. 2019;00:1–13. <https://doi.org/10.1080/23789689.2019.1600960>.
- Feng S, Hategeka C, Grépin KA. Addressing missing values in routine health information system data: an evaluation of imputation methods using data from the Democratic Republic of the Congo during the COVID-19 pandemic. *Popul Health Metr*. 2021;19:1–14. <https://doi.org/10.1186/s12963-021-00274-z>.
- Liu Y, De A. Multiple imputation by fully conditional specification for dealing with missing data in a large epidemiologic study. *Int J Stat Med Res*. 2015;4:287–95.
- Hopkins J, Narasimhan M, Aujla M, Silva R, Mandil A. The importance of insufficient national data on sexual and reproductive health and rights in international databases. *eClinicalMedicine*. 2024. <https://doi.org/10.1016/j.eclinm.2024.102554>.
- Donders ART, Van Der HGJMG, Stijnen T, Moons KGM. Review: A gentle introduction to imputation of missing values. 2006;59:1087–91.
- Graham JW. Missing data analysis: making it work in the real world. *Annu Rev Psychol*. 2009;60:549–76.
- Young R, Johnson DR. Handling missing values in longitudinal panel data with multiple imputation. *J Marriage Fam*. 2015;77:277–94. <https://doi.org/10.1111/jomf.12144>.
- van Buuren S. *Flexible imputation of missing data*. 2nd ed. Boca Raton, Florida: CRC Press; 2018.
- Blazek K, van Zwieten A, Saglimbene V, Teixeira-Pinto A. A practical guide to multiple imputation of missing data in nephrology. *Kidney Int*. 2021;99:68–74. <https://doi.org/10.1016/j.kint.2020.07.035>.
- Madley-dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. *J Clin Epidemiol*. 2019;110:63–73. <https://doi.org/10.1016/j.jclinepi.2019.02.016>.
- WHO. 2018 Global Reference List of 100 Core Health Indicators (plus health-related SDGs). Geneva; 2018. <https://www.who.int/healthinfo/indicators/2018/en/>.
- UNDP. *Human Development Report 2019: beyond income, beyond averages, beyond today*. 2019.
- Buuren S van. Long and wide format. In: *Flexible imputation of missing data*, 2nd edn. Boca Raton, Florida: CRC Press; 2018. p. 311–2.
- Schouten RM, Lugtig P, Vink G. Generating missing values for simulation purposes: a multivariate amputation procedure. *J Stat Comput Simul*. 2018;88:2909–30. <https://doi.org/10.1080/00949655.2018.1491577>.
- Fielding S, Fayers PM, Ramsay CR. Investigating the missing data mechanism in quality of life outcomes: a comparison of approaches. *Health Qual Life Outcomes*. 2009;7:57. <https://doi.org/10.1186/1477-7525-7-57>.
- Curnow E, Cornish RP, Heron JE, Carpenter JR, Tilling K. Multiple imputation using auxiliary imputation variables that only predict missingness can increase bias due to data missing not at random. *BMC Med Res Methodol*. 2024;24:231. <https://doi.org/10.1186/s12874-024-02353-9>.
- Curnow E, Tilling K, Heron JE, Cornish RP, Carpenter JR. Multiple imputation of missing data under missing at random: including a collider as an auxiliary variable in the imputation model can induce bias. *Front Epidemiol*. 2023;3:1237447.
- van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45:1–67.
- Wulff JN, Ejlskov L. Multiple imputation by chained equations in praxis: guidelines and review. *Electron J Bus Res Methods*. 2017;15:41–56.
- Huque MH, Carlin JB, Simpson JA, Lee KJ. A comparison of multiple imputation methods for missing data in longitudinal studies. *BMC Med Res Methodol*. 2018;18:168. <https://doi.org/10.1186/s12874-018-0615-6>.

35. Twisk J, de Vente W. Attrition in longitudinal studies: how to deal with missing data. *J Clin Epidemiol*. 2002;55:329–37. [https://doi.org/10.1016/S0895-4356\(01\)00476-0](https://doi.org/10.1016/S0895-4356(01)00476-0).
36. Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res*. 1999;8:3–15. <https://doi.org/10.1177/096228029900800102>.
37. Fichman M, Cummings JN. Multiple imputation for missing data: Making the most of what you know. *Organ Res Methods*. 2003;6:282–308.
38. Kim H-Y. Statistical notes for clinical researchers: a one-way repeated measures ANOVA for data with repeated observations. *Restor Dent Endod*. 2015;40:91–5.
39. Lee S, Lee DK. What is the proper way to apply the multiple comparison test? *Korean J Anesthesiol*. 2018;71:353–60.
40. Engels JM, Diehr P. Imputation of missing longitudinal data: A comparison of methods. *J Clin Epidemiol*. 2003;56:968–76.
41. Kim KH, Kim KJ. Missing-data handling methods for Lifelogs-Based Wellness Index Estimation: comparative analysis with panel data. *JMIR Med Informatics*. 2020. <https://doi.org/10.2196/20597>.
42. Kambach S, Bruelheide H, Gerstner K, Gurevitch J, Beckmann M, Seppelt R. Consequences of multiple imputation of missing standard deviations and sample sizes in meta-analysis. *Ecol Evol*. 2020;10:11699–712.
43. Luo Y. Evaluating the state of the art in missing data imputation for clinical data. *Brief Bioinform*. 2022. <https://doi.org/10.1093/BIB/BBAB489>.
44. Hayati Rezvan P, Lee KJ, Simpson JA. The rise of multiple imputation: a review of the reporting and implementation of the method in medical research. *BMC Med Res Methodol*. 2015. <https://doi.org/10.1186/S12874-015-0022-1>.
45. Waljee AK, Mukherjee A, Singal AG, Zhang Y, Warren J, Balis U, et al. Comparison of imputation methods for missing laboratory data in medicine. 2013;:1–7.
46. Haji-maghsoudi S, Haghdoost A, Rastegari A, Baneshi MR. Influence of Pattern of Missing Data on Performance of Imputation Methods : An Example Using National Data on Drug Injection in Prisons. 2013;1:69–77.

Publisher's Note

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