



## Research Article

# Accuracy of Five Multiple Imputation Methods in Estimating Prevalence of Type 2 Diabetes based on STEPS Surveys

Hamid Heidarian Miri<sup>1,\*</sup>, Jafar Hassanzadeh<sup>2</sup>, Saeedeh Hajebi Khaniki<sup>1</sup>, Rahim Akrami<sup>3</sup>, Ehsan Baradaran Sirjani<sup>4</sup>

<sup>1</sup>Department of Epidemiology and Biostatistics, Social Determinants of Health Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

<sup>2</sup>Department of Epidemiology, Research Center for Health Sciences, School of Health, Shiraz University of Medical Sciences, Shiraz, Iran

<sup>3</sup>School of Medicine, Sabzevar University of Medical Sciences, Sabzevar, Iran

<sup>4</sup>Department of Statistics, Islamic Azad University of Mashhad, Mashhad, Iran

## ARTICLE INFO

### Article History

Received 21 April 2019

Accepted 03 December 2019

### Keywords

Multiple imputation  
nonresponse  
STEPS surveys

## ABSTRACT

**Background:** This study was aimed to evaluate five Multiple Imputation (MI) methods in the context of STEP-wise Approach to Surveillance (STEPS) surveys.

**Methods:** We selected a complete subsample of STEPS survey data set and devised an experimental design consisted of 45 states ( $3 \times 3 \times 5$ ), which differed by rate of simulated missing data, variable transformation, and MI method. In each state, the process of simulation of missing data and then MI were repeated 50 times. Evaluation was based on Relative Bias (RB) as well as five other measurements that were averaged over 50 repetitions.

**Results:** In estimation of mean, Predictive Mean Matching (PMM) and Multiple Imputation by Chained Equation (MICE) could compensate for the nonresponse bias. Ln and Box-Cox (BC) transformation should be applied when the nonresponse rate reaches 40% and 60%, respectively. In estimation of proportion, PMM, MICE, bootstrap expectation maximization algorithm (BEM), and linear regression accompanied by BC transformation could correct for the nonresponse bias. Our findings show that even with 60% of nonresponse rate some of the MI methods could satisfactorily result in estimates with negligible RB.

**Conclusion:** Decision on MI method and variable transformation should be taken with caution. It is not possible to regard one method as totally the worst or the best and each method could outperform the others if it is applied in its right situation. Even in a certain situation, one method could be the best in terms of validity but the other method could be the best in terms of precision.

© 2020 Atlantis Press International B.V.

This is an open access article distributed under the CC BY-NC 4.0 license (<http://creativecommons.org/licenses/by-nc/4.0/>).

## 1. INTRODUCTION

In many countries, noncommunicable disease risk factors survey, also known as STEP-wise Approach to Surveillance (STEPS), is regarded as the main source of information on prevalence of type 2 diabetes. General structure of STEPS as well as the questionnaire and analysis process has been devised by World Health Organization [1].

However, some participants do not give their blood sample. As a result, biochemical variables including variable Fasting Blood Glucose (FBG) will contain some amount of missing data (on average about 19% in Iran's STEPS surveys), which can damage the estimates both in terms of precision and validity. The degree to which missing data damage the results and the decision on how to deal with it depends mostly on the mechanisms through which missing data arise.

There are three mechanisms under which missing data are produced: first, Missing Completely at Random (MCAR), which produces missing data independent from values in other variables as well as the missing value itself. This kind of missing data reduce the sample size but do not bias the estimates. Second, Missing at Random (MAR) in which the probability of being missing is independent from its value after adjusting for other variables. Third, Missing not at Random (MNAR) in which being missing depends on its value after adjusting for other variables. It has been argued that under MAR or MNAR, there is a need for action because, besides reduction in sample size, the estimates will be biased [2].

There are a wide range of methods to deal with missing data among which Multiple Imputation (MI) has gotten much attention in all fields of research [2,3]. Its popularity is for being highly efficient and the ability to produce unbiased results under MAR as well as MCAR assumptions [4].

Multiple imputation, proposed by Rubin in 1987 and developed by others including Schafer [5], substitutes each missing value with more than one ( $M$ ) value and produces  $M$  sets of data that are

\*Corresponding author. Email: [HeidarianH@mums.ac.ir](mailto:HeidarianH@mums.ac.ir)

Data availability statement: The data that support the findings of this study are available from the corresponding author, Hamid Heidarian Miri, upon reasonable request.

complete but different regarding imputed values. Then, the target analysis is performed on each data set and finally the corresponding estimates are combined according to the Rubin's combination rule [2] as follows [Equation (1)–(4)]:

$$Q_M = \frac{1}{M} \cdot \sum_{i=1}^M Q_i \quad (1)$$

$$B = \frac{1}{M-1} \sum_{i=1}^M (Q_i - Q_M)^2 \quad (2)$$

$$U = \frac{1}{M} \sum_{i=1}^M U_i \quad (3)$$

$$T = U + \left(1 + \frac{1}{M}\right) B \quad (4)$$

where  $Q_M$ : overall point estimate,  $Q_i$ : point estimate of  $i$ th-imputed data set,  $M$ : number of imputation,  $B$ : between imputations variance,  $U$ : within imputation variance,  $T$ : total variance.

During years, MI has been developed through several methods to fit different settings because its performance depends on the structure of data set, scale and distribution of variables, correlation among involved variables, rate and mechanism of missing data [6].

So, in order to find which method is best suited to estimate prevalence of type 2 diabetes based on STEPS data, we need to evaluate the accuracy of most advanced MI methods in the context of STEPS survey. This evaluation can be further justified through following reasons:

First, most of the studies that evaluated and compared the accuracy of MI methods have not included a wide range of methods as they are now available in statistical software. Second, no study has yet been made on the basis of STEPS studies, possibly due to the challenge that MI will not result in proper imputation in data sets with complex structure [2,7]. Third, it is not clear what percent of non-responses correspond to a certain amount of bias in STEPS results, so even if we accept that one method has a kind of superiority over others, it is desirable to know how much of nonresponse bias can be compensated by that method.

Regarding above explanations, this study was carried out to evaluate the accuracy of five MI methods in estimating prevalence of type 2 diabetes based on STEPS surveys. These MI methods include Expectation–Maximization with Bootstrapping Algorithm (BEM), Multivariate Normal Regression (MVN), Linear Regression (LR), Multiple Imputation by Chained Equation (MICE), and Predictive Mean Matching (PMM).

## 2. MATERIALS AND METHODS

### 2.1. Data Source

We used the data set of STEPS survey conducted by Iran's Ministry of Health in 2007 in which a sample of 30,000 of Iranian people aged 15–64 years had been recruited. Then, a subsample of data set was drawn in a way that the structure of survey was preserved. For this purpose, seven out of 30 provinces that had the least amount of missing data in variable FBG were selected. The selected subsample

consisted of 5099 participants and 21 variables on demographic, behavioral, anthropometrical, biochemical, and sampling characteristics. The full explanation of the STEPS survey and its results are beyond the scope of this article and are available elsewhere [8–10].

### 2.2. Experimental Design

Experimental design of this study consisted of 45 states ( $3 \times 3 \times 5$ ), which differed by rate of simulated missing data (20%, 40%, and 60%), variable transformation [no transformation, natural logarithm (Ln), and Box–Cox (BC)], and MI method (BEM, MICE, MVN, LR, and PMM).

### 2.3. MI Procedures

In this study, five MI methods were considered as follows:

*Bootstrap expectation–maximization algorithm* that uses expectation maximization algorithm introduced by Dempster [11] to estimate the posterior distribution of incomplete data and then a bootstrap method to take draws from this posterior distribution, which is considered fast and robust. Since this procedure lets us to simultaneously impute more than one variable without distinction between dependent and independent variables, all biochemical variables were included in the imputation process.

Multiple imputation by chained equation, also known as switching regression or sequential regression multivariate imputation [12], is completely formed on conditional specification models. The assumption of multivariate normality for data can be ignored [13–15]. Like BEM, biochemical variables were included in the imputation process. Since each variable is imputed using its own imputation model, MICE can handle different variable types (continuous, binary, unordered categorical, and ordered categorical) [16].

Multivariate normal regression with which missing values are imputed by taking values from posterior distribution estimated by Markov Chain Monte Carlo algorithm [4,5]. It is based on the multivariate normality, but generally, it is robust to this assumption and still can provide valid estimates. With MVN approach, it is possible to consider several response variables but all predictors should be completely observed. For MVN and two next approached, we applied MI system implemented in Stata version 11.2 (StataCorp. 2009. Stata Statistical Software: Release 11, College Station, TX: StataCorp LP).

Univariate Linear Regression (UVR) that univariates a posterior predictive distribution derived from a normal linear regression model that is univariate. The imputation model includes an imputing variable as the response and some complete predictors.

Predictive mean matching first introduced by Rubin [17]. It reduces the bias in a data set by means of imputation by taking real values sampled from the data. It is a semi-parametric version of the UVR with the same technique, but instead of displacing missing values with predictions of the model, PMM takes draws randomly from observed values whose predictive means are in a specified distance of that of missing values that has been set in advance.

To eliminate skewness in FBS to make it consistent with normality assumption, we used Ln and BC transformations.

## 2.4. Simulation

Simulation was restricted to generating missing data in variable FBG. In fact, for each state of the experimental design, the five following steps repeated 50 times:

Step 1: Generating MAR data in variable FBG in a way that probability of becoming missing depended to sex and then age: for this purpose, variable  $P$  was created that took 1 for men, 2 for women aged 25–36, and 3 for others. Then, variable  $R$  was created that took values between 0 and 1 drawn randomly from a uniform distribution. Next variable FBG was sorted in ascending order of variables  $P$  and  $R$  and its first 1000, 2000, and 3000 observations were converted to missing for 19.6%, 39.2%, and 58.8% missing data rates that were rounded to 20%, 40%, and 60%, respectively.

Step 2: Calculation of FBG-related measurements including mean of FBG and proportion with  $\text{FBG} \geq 126 \text{ mg/dl}$  for prevalence of type 2 diabetes. This step reflects how estimations were affected by missing data before MI. In estimation process, survey setting and weights, as defined by Iran's STEPS survey team, were taken into account. There is enough material available for more detailed information on sampling, weighting, and survey setting of the Iran's STEPS survey [18].

Step 3: Using MI method: the five considered MI methods were MICE, BEM, MVN, LR, and PMM. The independent variables used in all the methods were demographic, behavioral, anthropometrical, biochemical, and survey sampling variables [19]. The survey sampling variables were included because the missing pattern might differ between strata or clusters. Finally, the number of imputations was set to 20 [20,21]. Because detailed information on the imputation process for all MI methods is widely available in the literature [5,6,13,22,23], full explanation on MI methods was excluded from the scope of this work.

Step 4: Recalculation of FBG-related measurements is the same as in Step 2, which is intended to represent how MI was able to reduce the effect of missing data on the estimations.

Step 5: Calculation of Fraction of Missing Information (FMI), Relative Efficiency (RE), and Relative Variance Increased (RVI). MI system of Stata was used to calculate these measurements based on the following Equations (5)–(8) [2,24,25]:

$$\text{RE} = \left( 1 + \frac{\text{FMI}}{M} \right)^{-1} \quad (5)$$

$$\text{RVI} = \frac{(1 + M^{-1})B}{U} \quad (6)$$

$$\lambda = \frac{\text{RVI} + \frac{2}{(V_{\text{large}} + 3)}}{\text{RVI} + 1} \quad (7)$$

$$V_{\text{large}} = (M - 1) \left( 1 + \frac{1}{\text{RVI}} \right)^2 \quad (8)$$

where  $M$ : number of imputation,  $V$ : degrees of freedom,  $B$ : between imputations variance,  $U$ : within imputation variance.

As Steps 2, 4, and 5 were completed, the calculated measurements were recorded in a separate data set. In the end, this data set consisted of 2250 rows (50 simulations for each of the 45 states) and 10 columns representing measurements of Steps 2, 4, 5, and state.

## 2.5. Statistical Software

Multiple imputation by chained equation was performed by Stata 11.2 through ice command [26]. MVN, LR, and PMM were performed through MI system of Stata version 11.2. BEM was performed by R-based AMELIA II package version 3.4.2 [23]. However, for all MI methods, Steps 2, 4, and 5 were done by Stata to make the calculations consistent across all states.

## 2.6. Statistical Analysis

Based on FBG-related measurements derived in Steps 2 and 4, Relative Bias (RB), Average Length (AL) of 95% confidence interval, and Coverage Rate (CR) were computed for before and after MI. Then, these measurements along with RE, FMI, and RVI were averaged out over 50 simulations of each state.

Relative bias is defined as true value minus estimated value divided by the true value. In this context, true value was obtained from complete data set while estimated value was obtained from data set with missing data before and after imputation.

Average length is the difference between upper and lower limits of an estimate's confidence interval and indicates the precision of the estimate. AL is preferred over standard error since it takes into account the implication that the estimate is based on  $t$ -student distribution [5].

Coverage rate is the proportion of estimates that their confidence intervals contain the true value.

## 3. RESULTS

Tables 1–6 show RB, CR, and AL for before and after imputation, as well as RE, FMI, and RVI for MI methods, transformation, and rate of missing data. In Tables 1–3, the target measurement is the mean of FBG, and in Tables 3–6, the target measurement is proportion with  $\text{FBG} \geq 126 \text{ mg/dl}$ . For each target measurement, tables are separated by rate of missing data.

Table 1 shows that in estimation of mean of FBG when nonresponse rate is 20%, the PMM will result in the least RB ( $-0.01$ ) while LR with Ln transformation will have the best performance regarding AL (2.283), RE, FMI, and RVI. All methods except for MICE with LN transformation had perfect CR.

Table 2 demonstrates that in estimation of mean of FBG and when nonresponse rate reaches 40%, among methods with perfect CR, MICE will have the least RB (0.00) whereas BEM will have the narrowest AL (2.49). However, regarding RE, FMI, and RVI, the best performance belongs to PMM supplemented by BC transformation.

Table 3 displays that in estimation of mean of FBG and when non-response rate becomes 60%, PMM with BC transformation could result in the least RB (0.00) while BEM will have the best performance regarding AL (2.717), RE, FMI, and RVI. Both methods are among those with perfect CR.

Table 4 reveals that in estimation of proportion with  $\text{FBG} \geq 126 \text{ mg/dl}$  when nonresponse rate is 20%, PMM with BC transformation has the least RB ( $-0.40$ ) and in addition to some of the other methods has the perfect CR, AL, and RE. However, those methods

**Table 1** | Accuracy of multiple imputation methods in estimation of mean of fasting blood glucose when nonresponse rate is 20%

Transformation	Method	RB	CR	AL	RE	FMI	RVI
–	No	–0.21	1.00	2.57	–	–	–
No	MICE	–0.06	1.00	2.583	0.991	0.18	0.22
	BEM	0.03	1.00	2.526	0.991	0.16	0.19
	MVN	–0.04	1.00	2.583	0.990	0.18	0.22
	LR	0.06	1.00	2.544	0.991	0.17	0.20
	PMM	–0.01	1.00	2.393	0.994	0.11	0.13
Natural logarithm	MICE	–0.04	1.00	2.332	0.995	0.08	0.09
	BEM	0.08	1.00	2.283	0.995	0.09	0.10
	MVN	–0.03	1.00	2.340	0.995	0.09	0.10
	LR	0.07	1.00	2.283	0.995	0.08	0.09
	PMM	–0.04	1.00	2.400	0.994	0.11	0.13
Box–Cox	MICE	2.21	0.98	2.437	0.993	0.13	0.15
	BEM	0.15	1.00	2.369	0.990	0.1	0.11
	MVN	0.10	1.00	2.445	0.994	0.11	0.13
	LR	0.12	1.00	2.465	0.993	0.12	0.14
	PMM	–0.01	1.00	2.394	0.994	0.11	0.12

MICE, multiple imputation by chained equation; MVN, multivariate normal regression; LR, linear regression; PMM, predictive mean matching; RB, relative bias; CR, coverage rate; AL, average length of 95% confidence interval; RE, relative efficiency; FMI, fraction of missing information; RVI, relative variance increased.

**Table 2** | Accuracy of multiple imputation methods in estimation of mean of fasting blood glucose when nonresponse rate is 40%

Transformation	Method	RB	CR	AL	RE	FMI	RVI
–	No	–0.46	1.00	2.78	–	–	–
No	MICE	0.06	1.00	3.337	0.975	0.52	1.08
	BEM	0.01	1.00	3.169	0.974	0.49	0.97
	MVN	–0.04	1.00	3.365	0.974	0.53	1.13
	LR	0.07	1.00	3.377	0.973	0.53	1.16
	PMM	–0.13	1.00	2.789	0.979	0.41	0.71
Natural logarithm	MICE	0.00	1.00	2.564	0.982	0.36	0.58
	BEM	0.04	1.00	2.491	0.983	0.34	0.53
	MVN	0.09	0.98	2.489	0.982	0.34	0.53
	LR	1.92	0.98	2.499	0.981	0.37	0.60
	PMM	–0.13	1.00	2.634	0.983	0.34	0.51
Box–Cox	MICE	0.05	1.00	3.063	0.982	0.35	0.54
	BEM	0.34	1.00	2.562	0.980	0.33	0.50
	MVN	0.16	1.00	2.723	0.982	0.35	0.54
	LR	0.17	1.00	2.663	0.983	0.33	0.50
	PMM	–0.12	1.00	2.587	0.984	0.31	0.45

have some minor superiority regarding FMI and RVI but at the cost of much greater RB.

**Table 5** illustrates that in estimation of proportion of observations with FBG  $\geq 126$  mg/dl when nonresponse rate soars to 40%, MICE with BC transformation yields the least RB (0.72) and perfect CR, although there are methods that, in addition to perfect CR, have the better performance regarding AL, RE, and FMI among which PMM with Ln transformation has the least RVI, too.

**Table 6** indicates that in estimation of proportion of observations with FBG  $\geq 126$  mg/dl when nonresponse rate rises to 60%, MVN with BC transformation has the least RB (–0.82) but at the cost of losing much of the CR and of very high RVI. However, some of the other methods have perfect CR as well as highest RE and lowest FMI among which LR with BC transformation has shorter AL and lower RVI.

**Table 3** | Accuracy of multiple imputation methods in estimation of mean of fasting blood glucose when nonresponse rate is 60%

Transformation	Method	RB	CR	AL	RE	FMI	RVI
–	No	–4.07	0.00	3.83	–	–	–
No	MICE	–0.25	1.00	6.632	0.950	0.84	5.38
	BEM	–3.29	0.00	3.791	0.970	0.57	1.38
	MVN	52.70	0.39	134.905	0.950	0.99	105.63
	LR	–0.25	1.00	6.510	0.950	0.84	5.07
	PMM	–0.68	1.00	4.705	0.960	0.78	3.64
Natural logarithm	MICE	–0.17	1.00	4.150	0.960	0.76	3.28
	BEM	–2.47	1.00	2.717	0.970	0.47	0.88
	MVN	23.37	0.50	13.279	0.970	0.61	34.75
	LR	–0.25	1.00	4.199	0.960	0.76	3.24
	PMM	–0.28	1.00	4.107	0.960	0.72	2.56
Box–Cox	MICE	0.58	1.00	3.545	0.960	0.64	1.84
	BEM	–1.23	0.92	3.011	0.970	0.49	0.97
	MVN	21.52	0.49	14.708	0.970	0.60	26.19
	LR	0.71	1.00	3.514	0.960	0.63	1.72
	PMM	0.00	1.00	3.740	0.960	0.65	1.94

**Table 4** | Accuracy of multiple imputation methods in estimation of proportion of observations with fasting blood glucose  $\geq 126$  mg/dl when nonresponse rate is 20%

Transformation	Method	RB	CR	AL	RE	FMI	RVI
–	No	–3.81	1.00	0.017	–	–	–
No	MICE	–27.36	0.10	0.020	0.980	0.25	0.33
	BEM	–24.66	0.11	0.021	0.980	0.32	0.50
	MVN	–27.39	0.07	0.020	0.980	0.24	0.32
	LR	–25.66	0.05	0.019	0.980	0.23	0.30
	PMM	–1.08	1.00	0.016	0.990	0.12	0.14
Natural logarithm	MICE	–3.79	1.00	0.016	0.990	0.14	0.16
	BEM	–4.14	1.00	0.016	0.990	0.15	0.18
	MVN	–3.83	1.00	0.016	0.990	0.13	0.15
	LR	–2.79	1.00	0.016	0.990	0.13	0.15
	PMM	0.59	1.00	0.016	0.990	0.12	0.14
Box–Cox	MICE	1.51	1.00	0.016	0.990	0.13	0.15
	BEM	1.45	1.00	0.016	0.990	0.11	0.13
	MVN	1.70	1.00	0.016	0.990	0.11	0.13
	LR	1.94	1.00	0.016	0.990	0.12	0.13
	PMM	–0.40	1.00	0.016	0.990	0.13	0.15

## 4. DISCUSSION

In this study, five MI methods were evaluated quantitatively in the context of STEPS survey. In addition, the impact of variable transformation as well as the rate of missing data on the performance of the MI methods were examined. As an example, with 60% of nonresponse rate, the RB in proportion with FBG  $\geq 126$  mg/dl was –54.85% (highest amount in this study), CR was 0, and the length of CI was 0.03; however, MICE with BC transformation reduced the RB to 1.78%, increased the CR to 1, and reduced the length of CI to 0.02.

In fact, we found that it is not possible to regard one method as totally the worst or the best. Each method could outperform others if it is applied in its right situation. Even it is possible that in a certain situation, one method is the best in terms of validity but the other method is the best in terms of precision. So, findings of this study can serve as a guide for choosing the best strategy in case of

**Table 5** | Accuracy of multiple imputation methods in estimation of proportion of observations with fasting blood glucose  $\geq 126$  mg/dl when nonresponse rate is 40%

Transformation	Method	RB	CR	AL	RE	FMI	RFI
No	No	-8.81	1.00	0.019	—	—	—
	MICE	-55.35	0.00	0.026	0.970	0.47	0.90
	BEM	-52.20	0.00	0.031	0.960	0.62	1.69
	MVN	-57.42	0.00	0.026	0.970	0.47	0.91
	LR	-55.51	0.00	0.026	0.970	0.45	0.85
Natural logarithm	PMM	-2.62	1.00	0.018	0.980	0.29	0.42
	MICE	-10.24	0.98	0.019	0.980	0.31	0.45
	BEM	-9.98	0.98	0.019	0.980	0.34	0.53
	MVN	-10.11	0.96	0.018	0.980	0.30	0.43
	LR	-10.20	1.00	0.018	0.980	0.28	0.39
Box-Cox	PMM	-1.21	1.00	0.017	0.980	0.26	0.34
	MICE	0.72	1.00	0.018	0.980	0.31	0.46
	BEM	2.93	1.00	0.017	0.980	0.28	0.38
	MVN	1.22	1.00	0.017	0.980	0.29	0.42
	LR	1.26	1.00	0.017	0.980	0.26	0.35
	PMM	-2.69	1.00	0.017	0.980	0.27	0.38

**Table 6** | Accuracy of multiple imputation methods in estimation of proportion of observations with fasting blood glucose  $\geq 126$  mg/dl when nonresponse rate is 60%

Transformation	Method	RB	CR	AL	RE	FMI	RFI
No	No	-54.85	0.00	0.03	—	—	—
	MICE	-117.63	0.00	0.047	0.960	0.73	2.88
	BEM	-146.79	0.00	0.050	0.960	0.74	2.92
	MVN	-56.14	0.92	0.230	0.960	0.65	25.72
	LR	-119.27	0.00	0.045	0.960	0.72	2.61
Natural logarithm	PMM	-11.19	1.00	0.025	0.970	0.58	1.43
	MICE	-27.33	0.44	0.027	0.970	0.55	1.29
	BEM	-48.07	0.00	0.029	0.970	0.58	1.44
	MVN	-15.48	0.50	0.075	0.970	0.59	15.77
	LR	-28.93	0.42	0.027	0.970	0.57	1.42
Box-Cox	PMM	-7.74	1.00	0.022	0.970	0.50	1.00
	MICE	1.78	1.00	0.020	0.970	0.46	0.88
	BEM	-12.83	1.00	0.021	0.970	0.44	0.81
	MVN	-0.82	0.45	0.112	0.970	0.53	22.19
	LR	2.66	1.00	0.019	0.970	0.44	0.79
	PMM	-6.15	1.00	0.021	0.970	0.48	0.96

dealing with nonresponses in surveys like STEPS. However, if the performance of MI methods, as shown here, is dissatisfying, it is still possible to improve it. For this purpose, one way is to enhance the survey questionnaire with variable correlated with the target variable or to gather some information about unit nonparticipants [27,28].

Some features of the present study could be regarded as strength points. For example, for the first time, performance of a wide range of the most advanced MI methods, available in statistical software, was evaluated in the context of STEPS survey. Besides, simulation technique was used to generate missing data in a real data set, so the true value of missing data was known for evaluation of MI methods.

On the other hand, some limitations of this study should be considered. First, only one quantitative variable was used for generating missing data, imputation, and all estimations. Second, only three

rates of missing data were regarded, so the results remain silence for rates outside or between the ranges. Third, the number of simulations was arbitrarily set to 50 that might not be sufficient in some cases. Although the data set used in this study was not new, it does not impose a serious challenge as this study was not aimed to provide neither descriptive nor analytical measures but to clarify how to deal with nonresponses in STEPS surveys, which are almost the same regarding the features that affect the performance of MI methods.

In this study, the effect of number of variables on the performance of MI models was not assessed because of the correlation among variables [22,29].

Consistent with similar studies, MI methods were evaluated based on RB, CR, AL, RE, FMI, and RFI [6,30]. RB is a measure of invalidity and should be considered large if is more than 2% or 5% in either direction [6,31]. CR is also a measure of validity but at the same time is a function of AL, which is a measure of precision. RE, FMI, and RFI are a function of the number of imputation and the ratio of between to within imputation variance. Better imputation results in lower values of RFI and FMI while higher values of RE [2,24,25].

We observed that RFI differed slightly over five MI methods of each experimental state. However, in the state with 60% missing data, MVN had considerably higher RFI (Tables 3 and 6). This can be explained by the process we employed to generate missing data, which increased variance of FBG [31]. It seems that when rate of missing data reaches 60%, the increase in the variance cannot be tolerated by MVN.

Our findings show that REs were almost similar and ranged from 0.96 to 0.995 that was seen between states, which were different in other measurements. For example, PMM in Table 3 and BEM in Table 5 have the same RE as 0.96 but different RB as 0.68 and -52.2 and CR as 1.00 and 0, respectively. So, MI methods with the same RE should not necessarily be regarded completely equivalent.

It is said that FMI would be substantially less than the rate of missing data when the imputed data are highly predictive for the missing data [29] but we observe the contrary in some occasions. For example, in Table 4, RB is 26.31 while FMI is 0.14, almost 30% less than the rate of missing data. On the other hand, in Table 3, while RB is -0.68, FMI is 0.784, 30% more than the rate of missing data.

Our results showed that performance of MI methods was better when the target measurement was mean rather than proportion. For example, under MAR, 20% missing values could bias the proportion with  $FBG \geq 126$  mg/dl, while 60% missing values was needed to bias the mean of FBG.

In situations where missing values do not bias the estimates, decision on use of MI should be taken with caution as some MI methods can yield lower CR while some other ones can result in narrower confidence interval.

So, our findings are consistent with the literature indicating that performance of MI methods depends on the target analysis and rate of missing data and can be improved by selection of proper variable transformation [6]. Besides, as some studies have also shown in other settings, there are some MI methods with almost similar performance that can considerably correct for nonresponse bias in STEPS surveys [4,18,21,22].

## CONFLICTS OF INTEREST

The authors declare they have no conflicts of interest.

## AUTHORS' CONTRIBUTION

HH and JH contributed to the conceptualization and writing of the first draft of the manuscript. EB and RA contributed to literature research. HH and SH did data analysis, writing and revision of the manuscript. JH supervised investigation and methodology. HH supervised the research project. HH and SH reviewed and edited the final draft.

## FUNDING

This study was financially supported by Shiraz University of Medical Sciences grant number 92-6485.

## REFERENCES

- [1] World Health Organization (WHO). STEPwise approach to surveillance (STEPS). Geneva: WHO; 2014. Available from: <http://www.who.int/chp/steps/en/> [July 26, 2014].
- [2] Rubin DB. Multiple imputation for nonresponse in surveys. New York: John Wiley & Sons Inc.; 2004.
- [3] Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338:b2393.
- [4] Lee KJ, Carlin JB. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *Am J Epidemiol* 2010;171:624–32.
- [5] Schafer JL. Analysis of incomplete multivariate data. London: Chapman & Hall/CRC Press; 1997.
- [6] Qi L, Wang YF, He Y. A comparison of multiple imputation and fully augmented weighted estimators for Cox regression with missing covariates. *Stat Med* 2010;29:2592–604.
- [7] Binder DA, Sun W. Frequency valid multiple imputation for surveys with a complex design. In: Proceedings of the Survey Research Methods Section. Alexandria, VA: American Statistical Association; 1996, pp. 281–6.
- [8] Asgari F, Mirzazadeh A, Heidarian H. 3rd National Surveillance of Risk Factors of Non-Communicable Diseases (SuRFNCD-2007) in Iran. Iran: Ministry of Health and Medical Education; 2009, pp. 20–40.
- [9] Esteghamati A, Gouya MM, Abbasi M, Delavari A, Alikhani S, Alaeddini F, et al. Prevalence of diabetes and impaired fasting glucose in the adult population of Iran: National Survey of Risk Factors for Non-communicable Diseases of Iran. *Diabetes Care* 2008;31:96–8.
- [10] Asgari F, Aghajani H, Haghazali M, Heidarian H. Non-communicable diseases risk factors surveillance in Iran. *Iran J Public Health* 2009;38:119–22.
- [11] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B (Methodological)* 1977;39:1–38.
- [12] Kennickell AB. Imputation of the 1989 survey of consumer finances: stochastic relaxation and multiple imputation. In: Proceedings of the Section on Survey Research Methods. Atlanta, Georgia: American Statistical Association; 1991.
- [13] van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* 1999;18:681–94.
- [14] Raghunathan TE, Siscovick DS. A multiple-imputation analysis of a case-control study of the risk of primary cardiac arrest among pharmacologically treated hypertensives. *Appl Stat* 1996;45:335–52.
- [15] Raghunathan TE, Lepkowski JM, Van Hoewyk J, Solenberger P. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Surv Methodol* 2001;27:85–95.
- [16] White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 2011;30:377–99.
- [17] Rubin DB. Statistical matching using file concatenation with adjusted weights and multiple imputations. *J Business Econ Stat* 1986;4:87–94.
- [18] Heidarian Miri H, Hassanzadeh J, Rajaeefard A, Mirmohammakhani M, Ahmadi Angali K. Multiple imputation to correct for nonresponse bias: application in non-communicable disease risk factors survey. *Glob J Health Sci* 2016;8:133–42.
- [19] Esteghamati A, Meysamie A, Khalilzadeh O, Rashidi A, Haghazali M, Asgari F, et al. Third national Surveillance of Risk Factors of Non-Communicable Diseases (SuRFNCD-2007) in Iran: methods and results on prevalence of diabetes, hypertension, obesity, central obesity, and dyslipidemia. *BMC Public Health* 2009;9:167.
- [20] Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res* 1999;8:3–15.
- [21] Taylor JMG, Cooper KL, Wei JT, Sarma AV, Raghunathan TE, Heeringa SG. Use of multiple imputation to correct for nonresponse bias in a survey of urologic symptoms among African-American men. *Am J Epidemiol* 2002;156:774–82.
- [22] Lin TH. A comparison of multiple imputation with EM algorithm and MCMC method for quality of life missing data. *Qual Quant* 2010;44:277–87.
- [23] Honaker J, King G, Blackwell M. Amelia II: a program for missing data. *J Stat Softw* 2011;45:1–47.
- [24] Bernaards CA, Farmer MM, Qi K, Dulai GS, Ganz PA, Kahn KL. Comparison of two multiple imputation procedures in a cancer screening survey. *J Data Sci* 2003;1:293–312.
- [25] StataCorp LP. Stata multiple-imputation reference manual. College Station, TX: Stata Press. Accessed at 2013.
- [26] Royston P, White IR. Multiple imputation by chained equations (MICE): implementation in Stata. *J Stat Softw* 2011;45:1–20.
- [27] Kmetec A, Joseph L, Berger C, Tenenhouse A. Multiple imputation to account for missing data in a survey: estimating the prevalence of osteoporosis. *Epidemiology* 2002;13:437–44.
- [28] Moodie EEM, Delaney JAC, Lefebvre G, Platt RW. Missing confounding data in marginal structural models: a comparison of inverse probability weighting and multiple imputation. *Int J Biostat* 2008;4:1–23.
- [29] Lewis T, Goldberg E, Schenker N, Beresovsky V, Schappert S, Decker S, et al. The relative impacts of design effects and multiple imputation on variance estimates: a case study with the 2008 National Ambulatory Medical Care Survey. *J Offic Stat* 2014;30:147–161.
- [30] Carpenter JR, Kenward MG, Vansteelandt S. A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *J R Stat Soc Ser A* 2006;169:571–84.
- [31] Brockmeier LL, Kromrey JD, Hines CV. Systematically missing data and multiple regression analysis: an empirical comparison of deletion and imputation techniques. *Multiple Linear Regression Viewpoints* 1998;25:20–39.