

# The Effect of Multiple Imputations by Chained Equations on the Factors Associated with Immunization Coverage in India

Sachit Ganapathy<sup>1</sup>, Binukumar Bhaskarapillai<sup>2\*</sup>, Shailendra Dandge<sup>3</sup>

<sup>1</sup>PhD Scholar, Department of Biostatistics, JIPMER, Puducherry, India.

<sup>2</sup>Associate Professor, Department of Biostatistics, NIMHANS, Bengaluru, India

<sup>3</sup>Professor, Department of Pharmacology, MediCiti Institute of Medical Sciences, Hyderabad, India.

\*Corresponding Author: (Email: binukumarb@gmail.com)

## ABSTRACT

**Background:** National Family Health Survey-4 (NFHS-4) revealed a significant improvement in the percentage of complete immunization attained in India. Even though determinants of immunization coverage in India are addressed by some studies, the impact of missing data in such large-scale surveys has not been accounted earlier. The present study aimed to identify the potential factors associated with immunization coverage in India using the complete case analysis (CCA) and multiple imputations by chained equations (MICE) analysis.

**Materials and methods:** We created a dichotomous immunization variable based on the status of all the vaccines given to the child. All relevant variables were summarized using appropriate descriptive statistics along with the proportion of missingness. Further, MICE procedure was performed to impute the missing values after assessing the missing data mechanism. Multiple logistic regression after accounting for the sampling weights were used to report the estimates of odds-ratio (OR) and 95% confidence intervals (CI) for both CCA and MICE analysis and compared.

**Results:** The percentage of children under five years of age who had total immunization was 69%. Further, we observed that female sex and rural habitation had higher odds of getting immunized in both CCA and MICE. Moreover, wealth index, number of antenatal visits, checkup after delivery and place of birth played an important role in the immunization coverage.

**Conclusion:** MICE provided more precise risk estimates on potential factors associated with vaccination coverage compared to CCA, even if the major findings did not alter due to large sample size.

**Key Words:** Immunization, Health surveys, missing data, Logistic regression, complete case analysis, MICE

## INTRODUCTION

Immunization refers to the administration of a vaccine to protect a person from acquiring an infectious disease<sup>[1]</sup>. It is an effective and proven scheme in regulating and eradicating life-threatening infectious diseases. According to the World Health Organization (WHO), in the year 2018, nearly 19.4 million children did not receive complete immunization worldwide, and India was one amongst the top countries

with least coverage<sup>[2]</sup>. Although programmes like Universal Immunization Programme (UIP) and Mission Indradhanush, launched to strengthen the immunization coverage have done well, there is a need for introspection to improve the current scenario. Basic immunization includes vaccines provided within the first year of birth, and according to WHO guidelines, a child is fully immunized with all basic vaccinations if the child has

received Bacillus Calmette-Guerin vaccine against tuberculosis at birth, three doses each of Polio and pentavalent (diphtheria-tetanus pertussis-hepatitis B), Hemophilic influenza type B vaccines at 6, 10 and 14 weeks of age and a vaccination against measles at 9 months of age and three doses of Pneumococcal conjugate vaccine (PCV) given in three doses at 6, 10 and 14 weeks [3].

According to the National Family Health Survey (NFHS)-4, 62% children aged between 12-23 months received complete vaccination in India [4], an improvement from NFHS-3 report [5] which suggested that 43.4% of the children had received it. The NFHS-4 report also pointed out that Puducherry has the highest, 91.2% and Nagaland has the lowest, 35.4% vaccine coverage; whereas according to NFHS-3, Tamil Nadu had the highest, 80.9% and Nagaland had the lowest, 21% vaccine coverage.

Missing data of any form in research are an evident problem in health investigations. Missingness in the data not only decreases statistical power but also induces bias in the estimation of population parameters and the samples tend to be unrepresentative, in turn complicating the analysis of the study [6]. Each of these distortions may threaten the validity of the study and can lead to invalid conclusions. According to Strengthening the Reporting of Observational studies in Epidemiology (STROBE) guidelines [7], evidence has to be given on the appropriate technique used to deal with missing data and the distribution of missing information. Even the Consolidated Standards of Reporting Trials (CONSORT) guidelines [8] for clinical trials recommend a similar approach.

The traditional analysis, complete case analysis (CCA) fails to address the issue of missing data effectively if missingness is more than 10% as this approach loses power and adds to the bias that creeps in later with loss to follow-up. The guidelines specify that participants can

only be included in the analysis if their outcomes are imputed.

Multiple imputation technique is an extensively used technique in health surveys to deal with missing data. The theory of multiple imputation technique makes two major assumptions regarding the data of multivariate normality and data to be missing at random [9]. Further the multiple imputation technique consists of three major phases: the imputation phase, the analysis phase and the pooling phase [10].

The imputation phase makes several copies of the dataset, such that each dataset has different estimates for the missing data. Imputation phase consist of two steps; an imputation step (I-step) and a posterior step (P-step). The I-step predicts the missing values of the incomplete variables from the observed variables using a combination of regression equations built from estimates of mean vector and the covariance matrix. The imputed equation is expressed as,

$$Y_i^* = [\widehat{\beta}_0 + \widehat{\beta}_1(C_i)] + e_i$$

Where  $Y_i^*$  contains the regression coefficients that generate the predicted estimates and  $e_i$  is residual of the regression equation assumed to follow normal distribution. The purpose of the I-step is to create a collection of  $m$  datasets, where every dataset brings in unique set estimates of missing values. The P-step of the imputation phases estimates the mean vector and the covariance structure from the datasets created in the I-step. Subsequently it uses Monte Carlo simulation method to create fresh set of parameters by adding some residual terms to the mean vector and covariance matrix. Iteratively, these steps produce multiple copies of the dataset, each with different sets of estimates for the missing values.

The post-imputation phase follows the analysis phase, where each of the  $m$  dataset is considered for the required analysis individually. The analysis phase yields  $m$  different estimates of each dataset, which is then pooled into a single point estimate and is defined as,

$$\hat{\theta} = \frac{1}{m} \sum_{i=1}^m \hat{\theta}_i$$

where  $\hat{\theta}_i$  is the parameter estimate from each dataset and  $\hat{\theta}$  is the pooled estimate. Moreover, this phase also gives  $m$  estimates of standard errors and are then pooled using the Rubin's rule.

Multiple imputations by chained equations (MICE) offers advantage over other methods, to deal with missing data, in terms of flexibility of application in wide range of studies<sup>[11]</sup>. MICE impute data as a sequence of steps along with diagnostic checking following each step.

The present study aims to identify the factors associated with the level of immunization in India. Further the risk estimates of factors associated with immunization obtained from CCA and MICE were compared.

## MATERIALS AND METHODS

Data for the study was procured from the NFHS- 4 dataset, which comes in line as the fourth in the series of NFHS conducted by the Ministry of Health and Family Welfare, Government of India. The data is collected in the rural area through a two-stage sampling design with villages as the Primary Sampling Units (PSUs) at first stage and a random selection of 22 households in each PSU at second stage. In urban areas, Census Enumeration Blocks (CEB) were selected at the first stage and a random selection of 22 households in each CEB at the second stage.

The dataset was obtained from the *Demographic and Health Surveys (DHS)* website. The dataset chosen for the analysis was from the Individual's data (IAIR71FL). The factors related to immunization of the child were all available from a single dataset. The Women's dataset contained 4797 variables pertaining to socio-demographic, human immune virus (HIV) infection related, immunization, domestic violence, etc. and had a total of 6,99,686 subjects. A selected number of variables pertaining to immunization coverage were

selected based on scientific evidence. This analysis was restricted to women who had children less than five years of age and thereby, the sample size reduced to 1,90,898 subjects.

According to the definition of WHO for immunization, a binary outcome variable called immunization factor (IMF), was computed based on combination of all the vaccines available. In the dataset and if the child had received all the vaccines, it was coded as 'yes', otherwise as 'no'. Various variables considered for the immunization model were, IMF, sex of the child, type of place of residence, highest education level, wealth index, check up after delivery, delivery by Caesarean section, place of delivery, number of antenatal checkup, number of living children, age of the child in months, height of the child, weight of the child, hemoglobin level adjusted to altitude. In addition, number of checkups in first 10 days after delivery and person who did the checkup was also included as auxiliary variables in the model.

## Statistical Analysis

Distribution of the potential factors affecting immunization coverage was summarized using descriptive statistics. Median along with first and third quartile values were used to summarize non-normal quantitative variables; whereas frequency and percentage for qualitative variables. Distributional property of quantitative variables was tested using Kolmogorov-Smirnov test. Chi-square test was used to find the association between the qualitative variables; whereas Mann-Whitney U test was used to test the significance of difference in medians across dichotomous immunization coverage variable. Binary logistic regressions were performed to select the variables to be retained in the final model. The cut-off for the variable to be retained in the final model was kept at  $p < 0.2$  at the bivariate analysis. The unadjusted odds ratio (OR) and 95% confidence intervals (CI) were also obtained for the same. Further, multiple logistic regression

for the selected variables was performed by including all the relevant variables in the model. In multiple logistic regression analysis, the cut off for each variable was kept at  $p < 0.05$  to be validated as statistically significant in contributing for the immunization coverage. This complete case analysis was restricted to a sample of 148,067. As the NFHS- 4 survey adopted a multi-stage sampling, all estimates were obtained after accounting the probability weights. Moreover, a two-stage survey design was declared for the dataset.

As we observed a maximum amount of missingness of 20.7% for hemoglobin (Hb) among the variables pertaining to vaccination coverage, the method multiple imputation was used to overcome the loss of information due to missing value. The summary and the pattern of the missingness in the variables in the model were obtained to understand the missing data mechanism of the dataset. Further Little's MCAR test was used to test the missing data mechanism. For imputation, we adopted MICE approach after specifying the type of variables for the incomplete variables in the model. As the data augmentation failed to converge and the assumption of normality was violated for incomplete variables, chained equations were preferred over Markov Chain Monte Carlo (MCMC) technique. The duplicate dataset were created, to carry out the analysis, and then pool the estimates using Rubin's rule. The diagnostic plots were plotted to address the model fit in different completed dataset obtained from multiple imputation. After the imputation procedure, simple and multiple logistic regression were carried out. Moreover, the final model obtained from both complete case and multiple imputation analysis models were compared.

All statistical analysis was carried out using Statistical Package for Social Sciences (IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.) and Stata software (Stata Statistical Software: Release 13. College Station, TX: StataCorp LP).

## RESULT

This analysis includes a total of 190,898 women that had children less than five years of age among the 699,696 women who participated in the survey. Excluding, missing cases, 180,995 women had children under five years of age, out of which 124805(68.95%) had undergone total immunization.

We did observe significantly higher immunization coverage in female children as compared to male children, and there was also an observable difference between urban and rural population. We also found a trend in the proportion of immunization coverage in education level and wealth index, prompting the coverage to be lower amongst the wealthy and educated (Table 1).

In addition, we observed missingness in several variables including the dependent variable IMF. Among them, the variable 'hemoglobin level' had the highest missing percent with 20.65% and 'place of delivery' had the least missing percent with 0.05% (Table 1). Further, Little's MCAR test showed that the missing data might not be MCAR. Therefore, the MICE technique was used to impute the missing values from the list of variables having missingness.

Further, the simple and multiple logistic regression estimates after imputation showed no substantial changes in terms of risk estimates due to large sample size and small amount of missingness. However, the confidence intervals of the effect estimates for MICE were more precise compared to CCA. In the final model, all variables considered for the analysis found statistically significant association with immunization coverage in CCA (Table 2). However, it has been noticed that that female child has higher odds of being immunized as opposed to male child even though their proportion distribution between groups weren't significantly different. Both CCA and MICE analysis showed that rural population had greater immunization as compared to urban population. Apart from that, the type of

residence, wealth index, number of antenatal visits, checkup after delivery and place of birth played a significant role in the immunization coverage. Surprisingly, it has been noted higher the number of living children, the likelihood of immunization

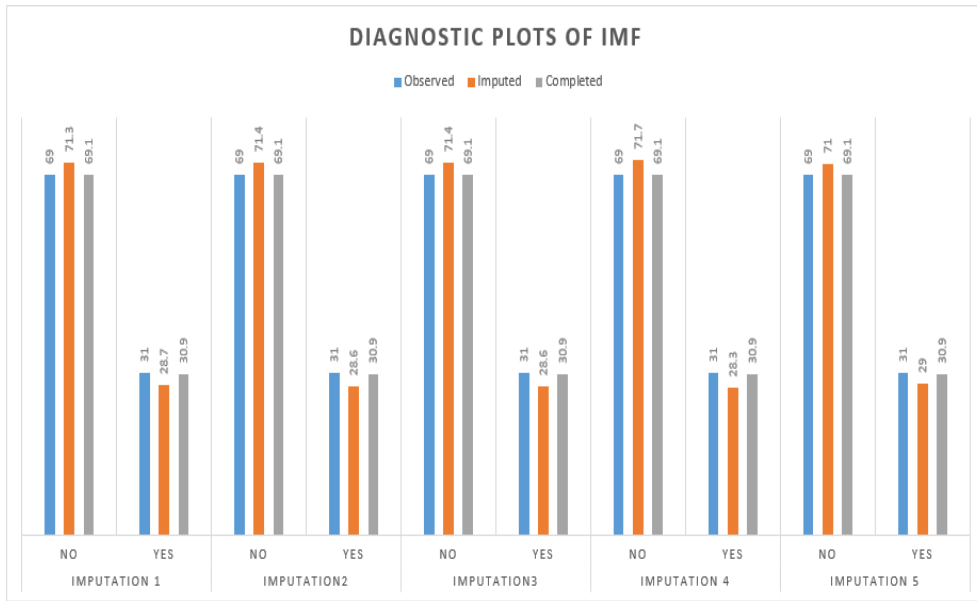
seemed to be relatively lower (OR: 0.92; 95%CI: 0.91, 0.93). The findings of CCA are in agreement with MICE analysis for all variables except 'age of the child'.

**Table 1: Distribution of the potential factors affecting immunization coverage**

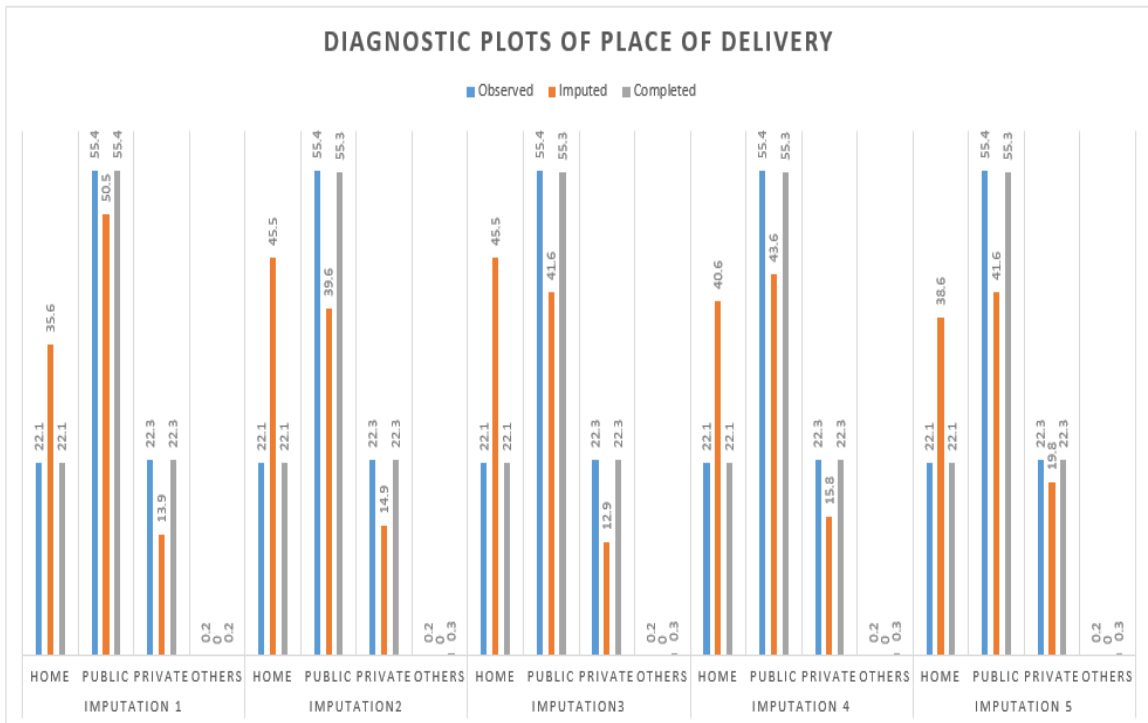
Variables		Immunization coverage		% Missing
		Yes (%)	No (%)	
Sex of the child	Male	67610 (68.9)	30506 (31.1)	0.0
	Female	57195 (69)	25684 (31)	
Type of Place of residence*	Urban	29856 (64.9)	16181 (35.1)	0.0
	Rural	94949 (70.4)	40009 (29.6)	
Education level*	Higher	12021 (61.2)	7614 (38.8)	0.0
	Secondary	55964 (65.7)	29249 (34.3)	
	Primary	18048 (72)	7016 (28)	
	No education	38772 (75.9)	12311 (24.1)	
Wealth Index*	Richest	16240 (58)	11741 (42)	0.0
	Richer	20500 (64.1)	11464 (35.9)	
	Middle	24871 (68)	11686 (32)	
	Poorer	29998 (73)	11105 (27)	
	Poorest	33196 (76.5)	10194 (23.5)	
Check up after delivery*	Yes	79615 (64.4)	44030 (35.6)	0.0
	No	45190 (78.8)	12160 (21.2)	
Delivery by C-section*	Yes	17394 (60.8)	11215 (39.2)	0.0
	No	107411 (70.5)	44975 (29.5)	
Place of delivery *	Others	315 (75.7)	101 (24.3)	0.05
	Private	27047 (66.2)	13783 (33.8)	
	Public	64644 (64.1)	36172 (35.9)	
	Home	32799 (84.2)	6134 (15.8)	
Total number of antenatal visits <sup>#</sup> *		4 (2,7)	3 (1,5)	0.97
Number of living children <sup>#</sup> *		2 (1,3)	2 (1,3)	0.0
Age of the child (in months) <sup>#</sup> *		27 (17,39)	20 (8,38)	3.98
Weight of the child (in kgs) <sup>#</sup> *		104 (87,125)	93 (70,120)	6.44
Height of the child (in cms) <sup>#</sup>		841 (760,926)	785 (680,905)	6.73
Hemoglobin level (g/dL) <sup>#</sup>		10.6 (9.6,11.6)	10.6 (9.6,11.6)	20.65

Further diagnostic plots compare the imputed values with the observed values, so that unlikely imputed values can be detected to check for the performance of the imputation. For continuous variables there are cumulative distribution functions and for categorical variables, bar graphs are plotted and provided in Figure 1 and Figure 2 respectively. The results shown that the observed and the imputed values were in accordance in terms of percentage distribution for the outcome variable IMF, whereas there was discrepancy in the distribution of observed and imputed values for 'place of delivery'. Nonetheless, for all the continuous variables that were imputed, the distribution of observed and imputed values was similar.

**Figure1: Diagnostic plots for categorical variables**



**Figure 1.1: Diagnostic plots for ‘IMF’**



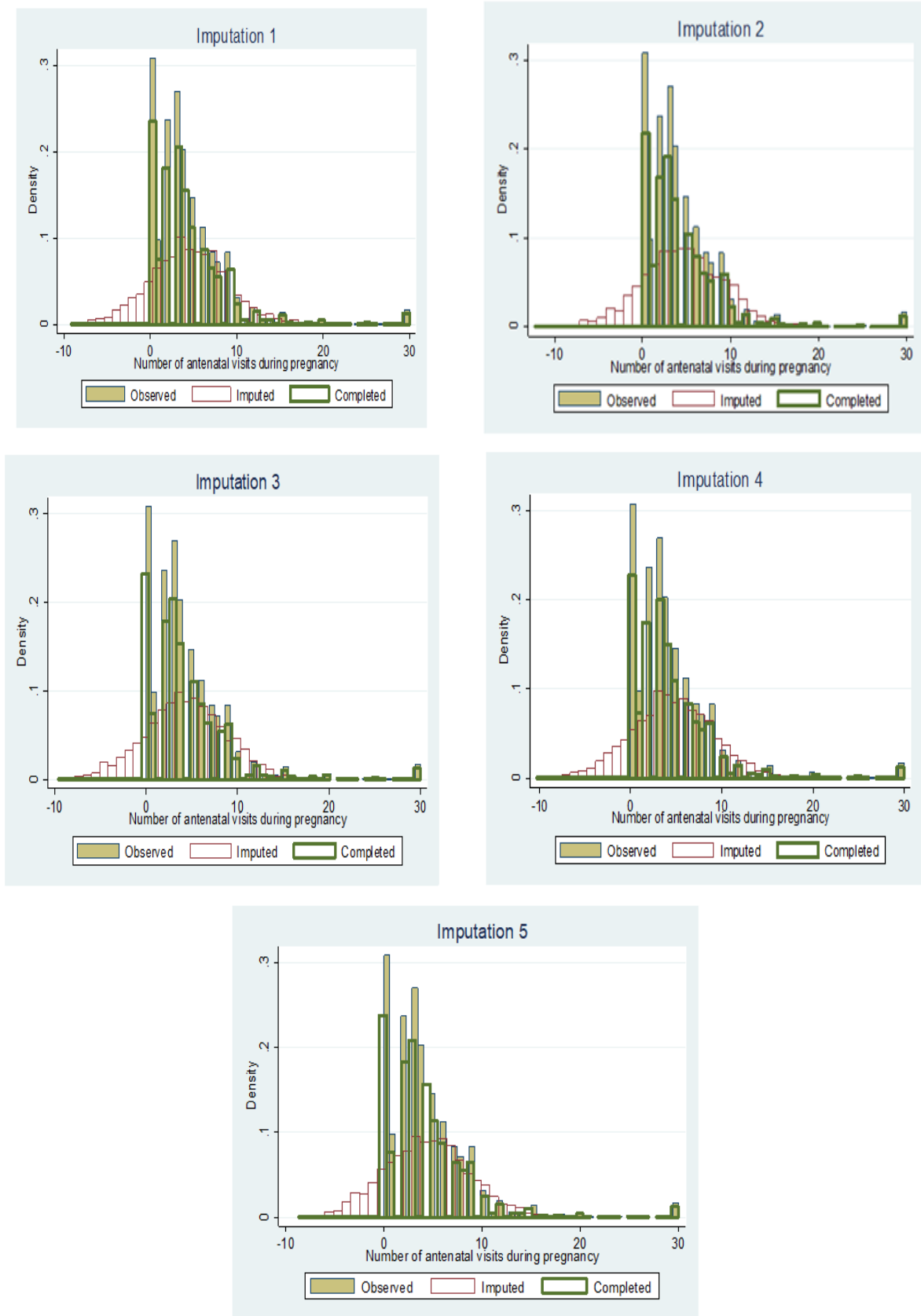
**Figure 1.2: Diagnostic plots for ‘place of delivery’**

**Table 2: Risk estimates of factors affecting the vaccination coverage by Complete Case and MICE analysis**

Variables		Complete case analysis				MICE analysis			
		Unadjusted OR (95% CI)	p- value	Adjusted OR (95% CI)	p-value	Unadjusted OR (95% CI)	p-value	Adjusted OR (95% CI)	p-value
Sex of the child	Female	1.02 (0.99, 1.05)	0.169	1.05 (1.03, 1.07)	<0.001	0.99 (0.97, 1.012)	0.461	1.07 (1.04, 1.09)	<0.001
Type of Place of residence	Rural	0.79 (0.76, 0.82)	<0.001	1.12 (1.09, 1.16)	<0.001	0.78 (0.76, 0.80)	<0.001	1.10 (1.07, 1.14)	<0.001
Education level	Higher	2.04 (1.94, 2.15)	<0.001	1.07 (1.02, 1.12)	0.007	1.98 (1.91, 2.05)	<0.001	1.01 (0.97, 1.06)	0.55
	Secondary	1.74 (1.68, 1.81)	<0.001	1.10 (1.06, 1.13)	<0.001	1.63 (1.59, 1.67)	<0.001	1.06 (1.03, 1.09)	<0.001
	Primary	1.29 (1.23, 1.36)	<0.001	1.05 (1.01, 1.09)	0.018	1.22 (1.18, 1.26)	<0.001	1.02 (0.98, 1.05)	0.33
Wealth Index	Richest	2.24 (2.14, 2.35)	<0.001	1.64 (1.57, 1.72)	<0.001	2.32 (2.25, 2.40)	<0.001	1.43 (1.37, 1.50)	<0.001
	Richer	1.82 (1.74, 1.91)	<0.001	1.23 (1.18, 1.28)	<0.001	1.80 (1.75, 1.86)	<0.001	1.14 (1.10, 1.19)	<0.001
	Middle	1.578 (1.512, 1.646)	<0.001	1.09 (1.05, 1.13)	<0.001	1.52 (1.47, 1.57)	<0.001	1.04 (1.01, 1.08)	0.021
	Poorer	1.27 (1.21, 1.32)	<0.001	0.98 (0.95, 1.02)	0.299	1.20 (1.16, 1.24)	<0.001	0.96 (0.93, 0.996)	0.025
Check up after delivery	Yes	1.85 (1.79, 1.92)	<0.001	1.46 (1.42, 1.50)	<0.001	2.02 (1.98, 2.07)	<0.001	1.39 (1.36, 1.43)	<0.001
Delivery by C-section	Yes	1.45 (1.39, 1.51)	<0.001	1.13 (1.09, 1.17)	<0.001	1.54 (1.50, 1.58)	<0.001	1.10 (1.06, 1.13)	<0.001
Place of delivery	Others	1.42 (1.06, 1.91)	<0.001	1.39 (1.09, 1.77)	<0.001	1.685 (1.34, 2.12)	<0.001	1.37 (1.08, 1.74)	0.009
	Private	2.35 (2.23, 2.47)	<0.001	1.45 (1.39, 1.51)	<0.001	2.71 (2.62, 2.80)	<0.001	1.51 (1.45, 1.57)	<0.001
	Public	2.53 (2.42, 2.65)	<0.001	2.19 (2.11, 2.27)	<0.001	2.98 (2.90, 3.07)	<0.001	2.21 (2.14, 2.29)	<0.001
Total number of antenatal visits		1.054 (1.05, 1.06)	<0.001	1.04 (1.037, 1.043)	<0.001	1.07 (1.066, 1.071)	<0.001	1.035 (1.03, 1.04)	<0.001
Number of living children		0.84 (0.83, 0.85)	<0.001	0.91 (0.90, 0.92)	<0.001	0.846 (0.84, 0.85)	<0.001	0.92 (0.91, 0.93)	<0.001
Age of the child (in months)		1.018 (1.017, 1.019)	<0.001	1.00 (0.998, 1.001)	0.749	1.019(1.019, 1.020)	<0.001	0.998(0.997, 1.00)	0.05
Weight of the child (in kgs)		1.011 (1.01, 1.011)	<0.001	0.998(0.997, 0.998)	<0.001	1.011(1.011, 1.012)	<0.001	1.001 (1.0, 1.001)	0.002
Height of the child (in cms)		1.002(1.002, 1.003)	<0.001	1.001(1.001, 1.001)	<0.001	1.003(1.002, 1.003)	<0.001	1.003(1.002, 1.003)	<0.001
Hemoglobin level (in g/dL)		0.999(0.999, 0.999)	<0.001	0.996(0.995, 0.997)	<0.001	1.006(1.005, 1.006)	<0.001	0.997(0.997, 0.998)	<0.001

Abbreviations: OR - Odds ratio, CI - Confidence Interval, kgs - kilograms, cms - centimeters, g-grams, dL-deciliters

**Figure 2: Diagnostic plots for continuous variables**



**Figure 2.1: Diagnostic plots for ‘Antenatal visit’**



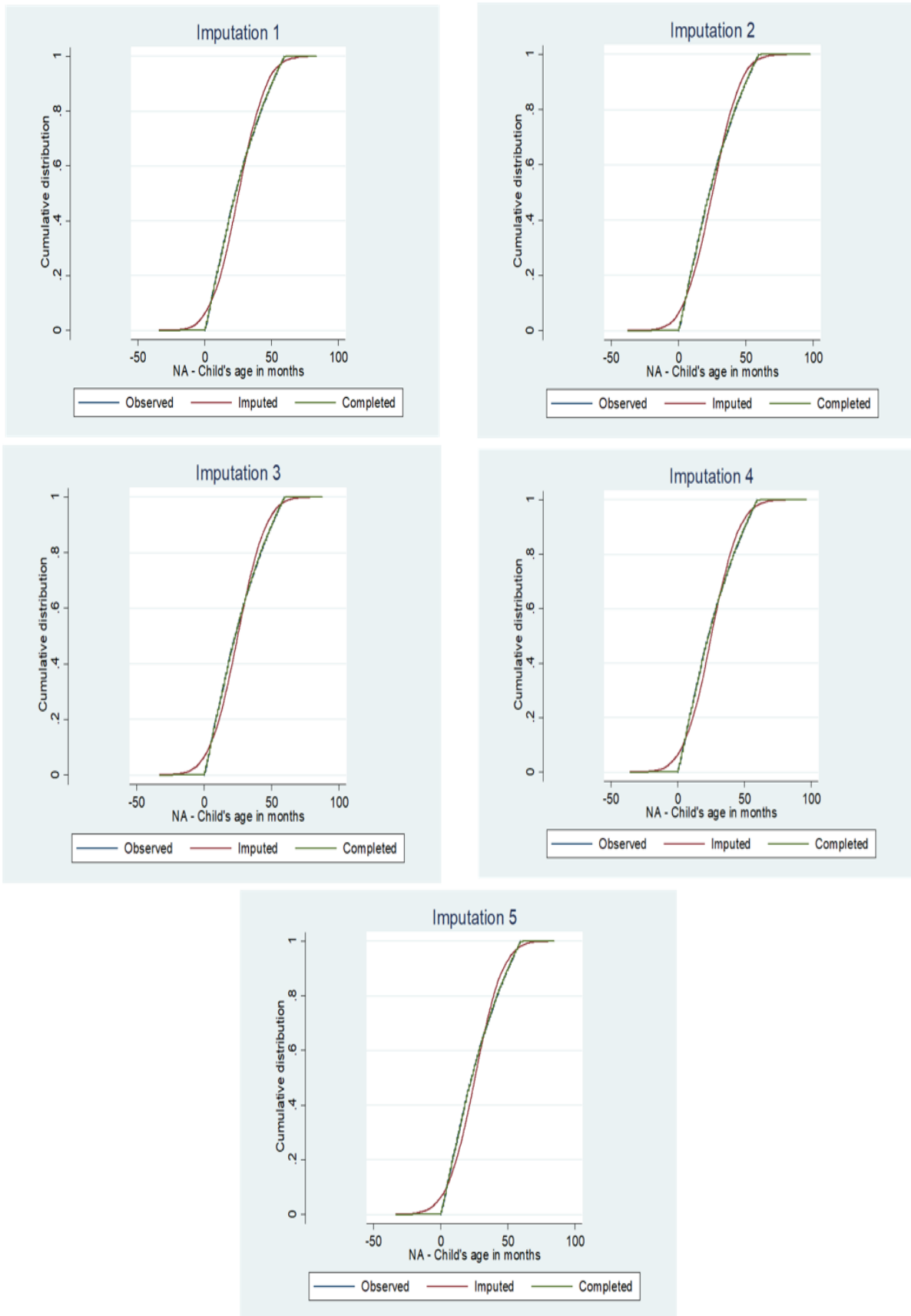


Figure 2.2: Diagnostic plots for 'age of child'

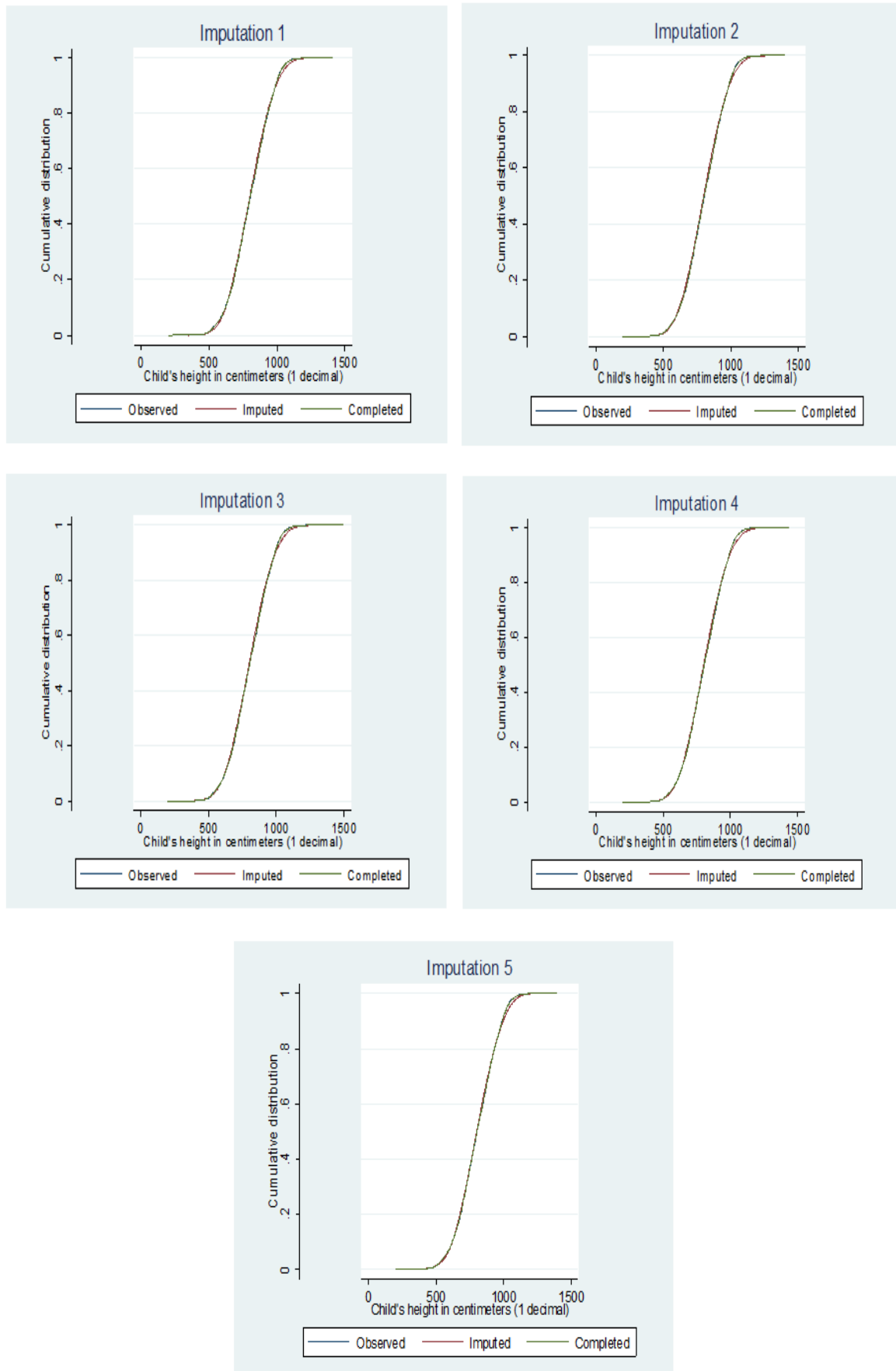


Figure 2.3: Diagnostic plots for 'height of child'

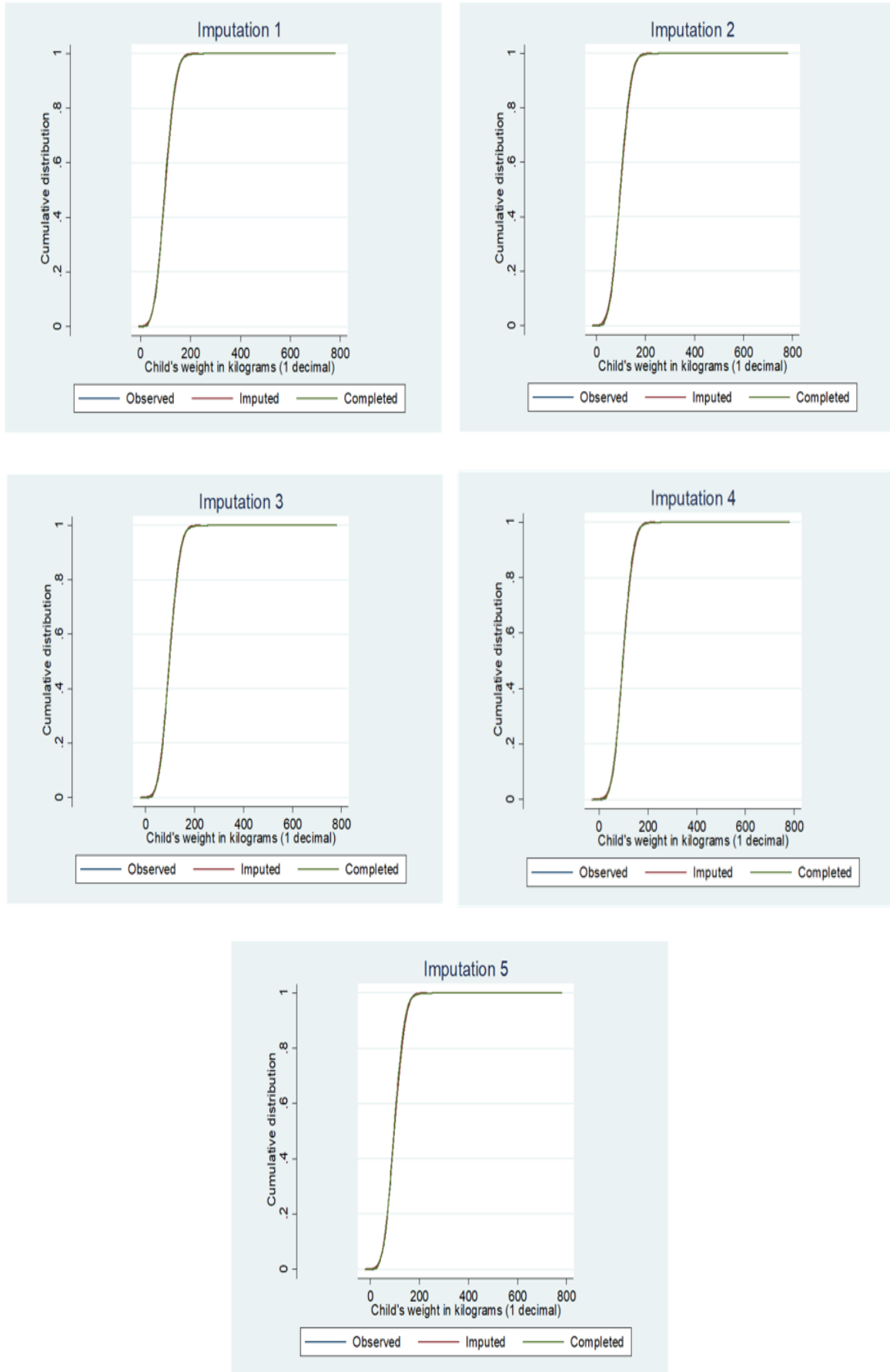


Figure 2.4: Diagnostic plots for 'weight of child'

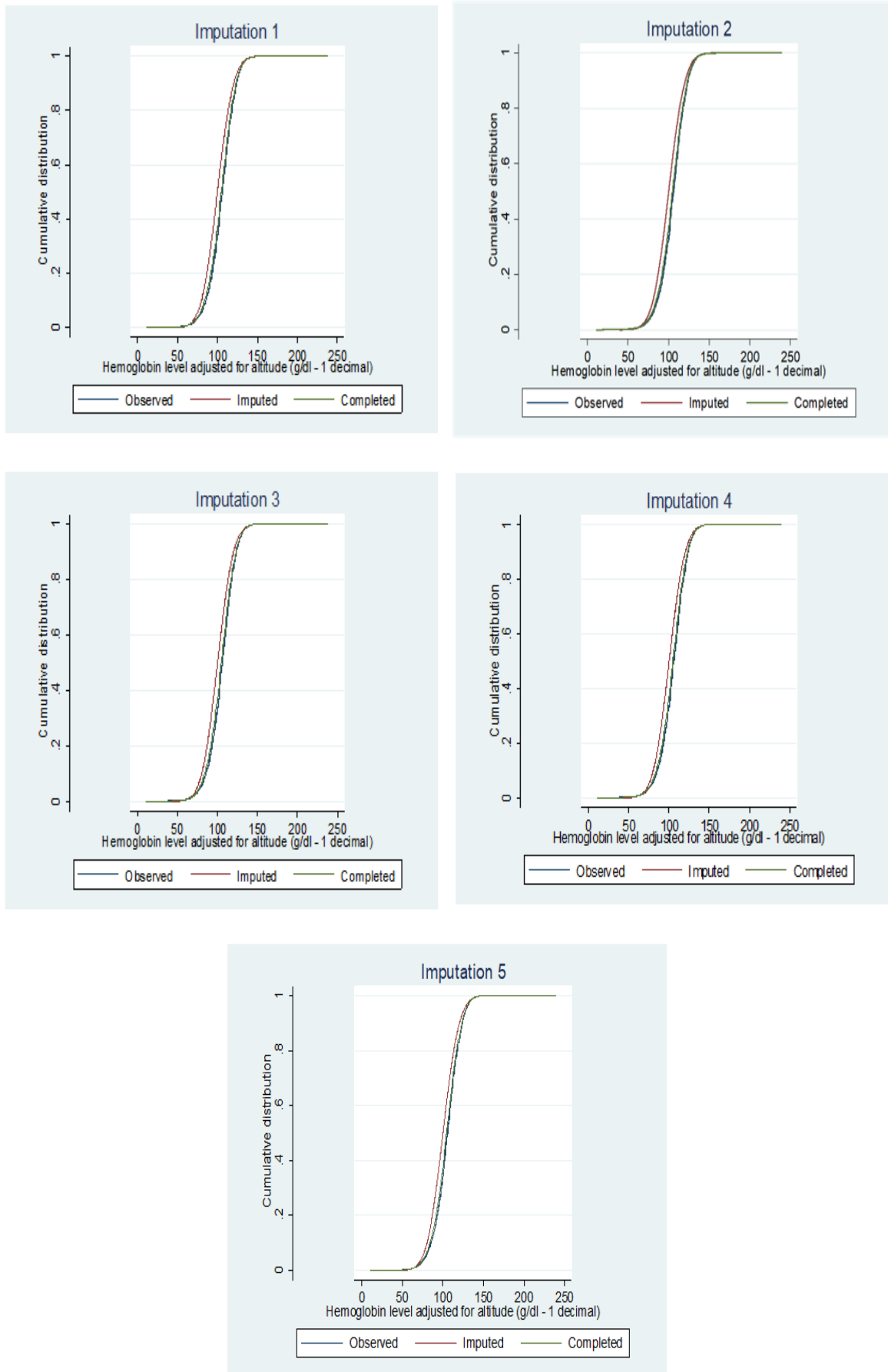


Figure 2.5: Diagnostic plots for 'hemoglobin level'

## DISCUSSION

In the current study, we implemented multiple imputation technique in dealing with missing data in assessing the factors affecting the immunization coverage in India from a nationally representative data from NFHS -4. We found that the percentage of missing observations ranged from 0.05% for *Place of delivery variable* to 20.7% for *Hemoglobin level (Hb variable)*. Overall, the percentage of missing data was 3% accounting for all the variables in consideration in the total sample size. Our analysis of determining the factors for immunization coverage accounted overall immunization coverage of 69% in both CCA and MICE methods of dealing with missing data. The important factors associated with low immunization coverage was *Sex of the child, Type of place of residence, Wealth index, Delivery by C-section, Check up after delivery and Place of delivery*. Other factors such as *Total number of antenatal visits and Number of children for the parents* also played an important in determining the immunization coverage. Both the CCA and MICE showed comparable results in handling missing data. Immunization has certainly increased over the past ten years among children between the ages of 12-23 months, from 43.5% according to NFHS – 3 to nearly 62% in NFHS – 4<sup>[12]</sup>.

According to the definition of WHO, and studies conducted subsequently have shown that post-natal care, place of delivery, wealth status, place of delivery, place of residence and number of living children were consistent factors which affected the immunization coverage<sup>[13]</sup>. In our study also all these variables were found to be significant contributors in the level of immunization. In addition, variables such as, sex of the child, type of delivery, age, weight, height and hemoglobin level of the child also played an important role in immunization coverage of the under-five aged children in India. A repeated cross-sectional study done to compare the results of all NFHS surveys among adolescent mothers reported that socioeconomic factors contributed to the disparity in immunization

coverage<sup>[14]</sup>. In large surveys like NFHS-4, which focuses on variables related to immunization, missing values are evident. MICE has proven to be an effective way to deal with missing data in literature, especially studies with larger proportion of missingness<sup>[15]</sup>. Adopting MICE techniques during data analysis guarantees that the representative sample used for the study reflect a true population. Most studies use a complete case (CC) analysis in order to overcome the issues of missing data. Single imputation method may not be valid under MAR condition, as it does not incorporate dependency that exists between the variables of interest<sup>[16]</sup>. However, multiple imputation technique addresses the problems of variability due to imputation, missingness in the dependent variable of regression analysis, the restriction on the missing data mechanism and distribution of the data of each variable<sup>[17]</sup>. Due to the large sample size in NFHS – 4, the effect of missing data may not be identified and acknowledged. But it is important to understand the factors which may play a significant role in immunization coverage, so that policies can be implemented effectively.

The strengths of the study are that analysis was performed using a large nationally representative sample. The model built to identify the factors affecting immunization coverage were all captured as explained from existing evidence. Moreover, MICE technique and CC analysis which are the two most commonly used methods for analyzing missing data were used. However, while performing MICE analysis, although studies suggested that number of imputations of 10 produce slightly better and reliable estimates, we restricted our analysis to five imputation datasets.

## CONCLUSION

Many guidelines, such as STROBE and CONSORT have highlighted the relevance of dealing with missing data and reporting it in research findings. However, this study re-emphasizes the need to account for the missing information in the data in order to obtain reliable risk estimates. In

conclusion, we observed that missing data, completely at random, is a common phenomenon in a large-scale health survey. In addition, MICE estimates were found to be more precise compared to CCA on the evaluation of potential risk factors associated with vaccination coverage, even if the major finding did not alter due to large sample size. Thus, we infer that MICE analysis can be an important alternative used and implemented in studies where missing data can cause aberrations to the results as compared to CCA.

**Contributors:** This study was conceived by the first (SG) and the second (BB) authors. Both SG and BB are equally responsible for manuscript preparation and statistical analysis. All the authors critically evaluated and contributed to the final version of the manuscript and approved for submission.

**Funding:** None

**Conflict of Interest:** None

## REFERENCES

1. Immunization WHO [Internet].2020. [Cited on 2020 July 15] Available from: <https://www.who.int/topics/immunization/en/>.
2. WHO Immunization coverage [Internet] 2020. [Cited on 2020 July 15] Available from: <https://www.who.int/news-room/fact-sheets/detail/immunization-coverage>.
3. World Health Organization, Immunization, vaccines, and biologicals: implementation research in immunization. 2019.
4. International Institute for Population Sciences. National Family Health Survey (NFHS-4), 2015-16: India. International Institute for Population Sciences; 2017
5. International Institute for Population Sciences. National Family Health Survey (NFHS-3), 2005-06: India. International Institute for Population Sciences; 2007.
6. Kang H. The prevention and handling of the missing data. Korean Journal of Anesthesiology. 2013 May 1;64(5):402-6.
7. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, Strobe Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. International Journal of Surgery. 2014 Dec 1;12(12):1495-9.
8. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. BMC Medicine. 2010 Dec;8(1):18.
9. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychological Methods. 2002 Jun;7(2):147.
10. Rubin D. B. (1987), "Multiple Imputation for Nonresponse in Surveys". John Wiley & Sons, Inc., New York
11. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? International Journal of Methods in Psychiatric Research. 2011 Mar 1;20(1):40-9.
12. National Family Health Survey-4 (2015-2016) Fact sheet. Mumbai: International Institute of Population Sciences; January 2016. 8 p.
13. Shrivastava N, Gillespie BW, Lepkowski JM, Boulton ML. Vaccination timeliness in children under India's universal immunization program. The Pediatric Infectious Disease journal. 2016 Sep 1;35(9):955-60.
14. Kumar C, Singh PK, Singh L, Rai RK. Socioeconomic disparities in coverage of full immunisation among children of adolescent mothers in India, 1990–2006: a repeated cross-sectional analysis. BMJ open. 2016 Aug 1;6(8)
15. Dohoo IR, Nielsen CR, Emanuelson U. Multiple imputation in veterinary epidemiological studies: a case study and simulation. Preventive veterinary medicine. 2016 Jul 1; 129:35-47.
16. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomized clinical trials—a practical guide with flowcharts. BMC medical research methodology. 2017 Dec;17(1):162.
17. McCleary L. Using multiple imputation for analysis of incomplete data in clinical research. Nursing Research. 2002 Sep 1;51(5):339-43.
18. Gurnani V, Haldar P, Aggarwal MK, Das MK, Chauhan A, Murray J, Arora NK, Jhalani M, Sudan P. Improving vaccination coverage in India: lessons from Intensified Mission Indradhanush, a cross-sectoral systems strengthening strategy. BMJ. 2018 Dec 7;363.
19. Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis. Applied Statistics. 1994 Jan 1:49-93.
20. Little RJ, Schenker N. Missing data. In Handbook of statistical modeling for the social and behavioral sciences 1995 (pp. 39-75). Springer, Boston, MA.
21. Kim KH, Bentler PM. Tests of homogeneity of means and covariance matrices for multivariate incomplete data. Psychometrika. 2002 Dec 1;67(4):609-23.
22. Rubin DB. Inference and missing data. Biometrika. 1976 Dec 1;63(3):581-92.

How to cite this article: Ganapathy S, Bhaskarapillai B, Dandge S. The effect of multiple imputations by chained equations on the factors associated with immunization coverage in India. *Int J Health Sci Res.* 2021; 11(6): 249-262. DOI: <https://doi.org/10.52403/ijhsr.20210638>

\*\*\*\*\*