


# The Use of Multiple Imputation to Handle Missing Data in Secondary Datasets: Suggested Approaches when Missing Data Results from the Survey Structure

INQUIRY: The Journal of Health Care  
Organization, Provision, and Financing  
Volume 59: 1–6  
© The Author(s) 2022  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/00469580221088627  
journals.sagepub.com/home/inq



Soojung Jo, PhD, RN<sup>1</sup> 

## Abstract

Secondary datasets are used in healthcare research because of its cost advantages, its convenience, and the size of the datasets. However, missing data can cause problems that are difficult to resolve. This manuscript reviews possible causes for missing data, and how to address them. Many researchers use multiple imputation as a solution, which consists of three phases: (a) the imputation phase, (b) the analysis phase, and (c) the pooling phase. When missing data is caused by a refusal to answer or by insufficient knowledge, multiple imputation works well. However, difficulties arise when there are problems with screening questions. If respondents do not answer a screening question, possible answers could be either “yes” or “no.” This paper suggests identifying “yes” responses on the screening question, and setting them aside for use in the analysis. The reasons for this approach are the impossibility of conducting multiple imputation twice, the problem of imputation based on the population after sample weight, and the difficulty of producing logical errors on the estimation in imputation phase. This manuscript uses as an example the techniques used to address missing data from screening questions in a national US dataset. These techniques of multiple imputation using examples from the dataset could be used by researchers in future healthcare research that relies on secondary datasets.

## Keywords

method, missing data, secondary analysis, multiple imputation, quantitative

### What do we already know about this topic?

- The three causes of missing data in secondary data are (a) survey structure, (b) refusal to answer, and (c) insufficient knowledge about questions.
- Inappropriate handling of missing data produces biased results.
- There is a lack of evidence surrounding handling missing data at the level of screening question.

### How does your research contribute to the field?

- This paper reviews handling missing data in a secondary dataset using the public secondary data for explaining different techniques for handling missing data.
- This paper suggests portioning out the cases answered “yes” in the screening question and using that portion of the dataset for analysis.

### What are your research’s implications towards theory, practice, or policy?

This manuscript would be helpful to conduct the future research encountering the missing data in the future secondary data analysis studies by explaining the steps of conducting multiple imputation.

## Background

Secondary datasets generally refer to large datasets collected by government or research institutions that contain a wide-ranging

sample of individuals and that usually represent the wider population.<sup>1</sup> Secondary datasets could be either quantitative or qualitative, and could include data obtained from surveys, interviews, or personal health records. The methods for analyzing datasets



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

differ based on the type of dataset.<sup>2</sup> For quantitative secondary datasets, it is common to conduct descriptive or inferential analysis.<sup>2</sup>

The most significant advantages of using secondary datasets are lower cost and higher efficiency so researchers can limit time spent designing, collecting, and organizing data.<sup>1,3</sup> In healthcare, a significant volume of research uses government-produced secondary datasets. Many researchers prefer to use these secondary datasets because they are usually free, easy to download, and generally include large samples. The data collection strategy of the datasets is developed by more than one expert, so the quality of data in the secondary dataset can be more rigorous than the primary dataset. Moreover, the data usually has been cleaned by the collecting organization and has been formatted for specific types of statistical software making it ready-to-use.<sup>3</sup> Instructional materials for researchers may be provided along with the data to explain survey items and to suggest analysis options, as well as providing programming code for different software packages. A final advantage of using secondary data is that it avoids problems associated with the need to repeat studies that address sensitive topics.<sup>4</sup>

However, secondary data also has disadvantages. Researchers are working with data collected through questions that they did not write, so the dataset survey design may not be relevant or the wording of the survey items may not exactly match the researcher's purpose.<sup>1</sup> Moreover, secondary datasets may not address particular problems of interest to the researcher, so some important variables may be missing from the analysis.<sup>3</sup> Finally, secondary data may be incomplete or contain problems associated with missing or implausible data.<sup>5</sup> Missing data is a significant obstacle to researchers engaged in the analysis of secondary datasets.

### Missing Data in Secondary Datasets

Frequently, secondary datasets have missing data.<sup>5</sup> Some of the reasons for missing data in cross-sectional secondary datasets are (a) survey structure, (b) refusal to answer, and (c) insufficient knowledge.<sup>6</sup> Some surveys have screening questions that are designed to collect answers from a particular population. For example, if the survey includes questions about cervical cancer screening, a percentage of questions may be relevant only for female respondents. In this situation, males will have missing data associated with cervical cancer screening questions.

A second reason for missing data are survey recipients who refuse to answer specific questions. If the survey has

personal questions, sensitive questions, or questions that cause respondents to feel uncomfortable, survey recipients may choose not to respond.<sup>7</sup> Previous literature indicates that people often refuse to answer personal or sensitive questions.<sup>7,8</sup>

A third reason for missing data is that survey recipients have insufficient knowledge to answer the questions. If a question is hard for survey recipients to understand, they often do not attempt to answer the question.<sup>7</sup> Even when a survey includes "don't know" answers, many recipients choose not to answer the question at all rather than answer "don't know".<sup>7</sup>

The approach a researcher uses to handle missing data can impact statistical results and study reliability.<sup>9</sup> Research has been conducted to specifically address the problems of missing data in secondary datasets. This study reviews possible ways to handle missing data in secondary datasets based on the causes of missing data, including missing data caused by the survey structure. The Health Information National Trends Survey (HINTS) 5 cycle 2<sup>10</sup> conducted in 2018 is used as an example to help readers understand the multiple imputation process. This manuscript addresses the multiple imputation process within the specific case of the analysis of the association between Human Papillomavirus (HPV) awareness and HPV vaccine recommendation.

### Methods

The first aim of this study is to review handling missing data in a secondary dataset. The literature review included research papers and books that addressed problems of missing data in a secondary dataset using search engines including PubMed. The search used keywords and terms including "missing data," "secondary data," and "multiple imputation." The second aim of this study is to use the HINTS dataset as the example for explaining different techniques for handling missing data. Anyone can download the HINTS on their website.

### Material

The HINTS was developed by the US National Cancer Institute to assess cancer-related information use and behaviors associated with cancer prevention. Annual data collection is based on a stratified random selection of mailing addresses. Mail with the survey questions is sent to potential participants, and responses are returned to the research center through the mail. The response rate was

---

<sup>1</sup>College of Nursing and Health Innovation, Arizona State University, Phoenix, AZ, USA

Received 18 August 2021; revised 19 February 2022; revised manuscript accepted 25 February 2022

#### Corresponding Author:

Soojung Jo, Edson College of Nursing and Health Innovation, Arizona State University, 550 N 3rd Street, Phoenix, Arizona 85004, USA.  
Email: soojungi@asu.edu

32.85% in the 2018 dataset and answers from 3,504 responses were published.

This manuscript uses two variables for discussion: HPV awareness and HPV vaccine awareness. HPV awareness was assessed by the question “Have you ever heard of HPV? HPV stands for Human Papillomavirus. It is not HIV, HSV, or herpes.” There were 50 missing responses (1.4%) for this question. HPV vaccine awareness was assessed by the question “A vaccine to prevent HPV infection is available and is called the HPV shot, cervical cancer vaccine, GARDASIL, or Cervarix. Before today, have you ever heard of the cervical cancer vaccine or HPV shot?” Neither HPV awareness nor HPV vaccine awareness had a screening question, so everyone could potentially answer the question. There were 92 missing responses (2.63%). If the records that included missing values in two variables were excluded (complete case analysis); 92 responses would have been excluded, resulting in a missing rate of 2.63%.

## Results

### *Handling Missing Data*

When a dataset has missing data, researchers need to understand the context in which the data is missing. There are three assumptions: missing completely at random, missing at random, and not missing at random.<sup>6</sup> In the assumption of missing completely at random, the missing data do not depend on either observed or non-observed variables.<sup>11</sup> In the missing at random condition, observed variables influence the missingness of the non-observed variable, but the non-observed variable does not itself contribute to the missing data.<sup>11</sup> That is, if the data is completely missing at random, it could also be missing at random. But, if the data is missing at random, then data may not always completely be missing at random. In the third possibility, not missing at random, both observed and not-observed variables have an influence on the missing data.<sup>6</sup>

For data that is missing completely at random, complete case analysis can be used. Complete case analysis, also known as Listwise deletion, is a statistical method for handling missing data. When analyzing data, complete case analysis excludes cases in which any variables are missing. In the condition of missing completely at random, the subset used in complete case analysis is the same subset used when researchers choose samples randomly in the dataset.<sup>11</sup> Eliminating cases which have missing data is convenient. Researchers do not need to restrict the statistical software, and the remaining dataset can be used for all analyses.<sup>12</sup> However, the assumption of missing completely at random is rarely made, because the condition is hard to test.<sup>11,12</sup>

When the assumption of the missing completely at random is not made, complete case analysis results in the distortion of parameter estimates.<sup>12</sup> Previous research has shown that the complete case analysis produces overestimated

means and inaccurate variabilities and correlations.<sup>12</sup> In the third condition of not missing at random, complete case analysis produces acceptable parameter estimates compared with other missing data handling methods if the percentage of missing values ranges from 5% to 10%.<sup>13,14</sup>

An additional concern associated with using complete case analysis is that the method is wasteful.<sup>12</sup> Even with a 10% missing rate in the larger dataset, there are likely to be sizable incomplete cases. Eliminating all cases with missing data reduces the sample size, which can impact the statistical power.<sup>12</sup> Because of all these limitations associated with complete case analysis, researcher began handling missing data using imputation.

### *Multiple Imputation*

Imputation is also a method for handling missing data. Single imputation is performed by replacing missing data with a single data value such as the mean.<sup>15</sup> Because single imputation does not take variability into consideration, the single imputation produces biased results, even in the missing completely at random condition.<sup>12</sup> To address this problem, researchers have explored multiple imputation.<sup>13</sup> Research has shown that using multiple imputation yields more accurate estimates of the missing data compared to complete case analysis with a large missing data rate.<sup>14,16</sup> Multiple imputation assumes missing at random, and consists of three phases: (a) the imputation phase, (b) the analysis phase, and (c) the pooling phase.<sup>12</sup>

### *Selecting Auxiliary Variables*

Before conducting multiple imputation, researchers need to decide on the variables related to dependent and independent variables that will be included in the imputation phase. This decision is important, because including unrelated variables or too many variables can cause a biased dataset.<sup>17</sup> For this example, the design is to run the regression using the HPV awareness as a dependent variable, and using HPV vaccine awareness, and demographic variables as the independent variables.

Auxiliary variables should be associated with the variables that include missing data, and can be found in the datasets based on the literature.<sup>18</sup> In this case, a literature review identified variables that are related to the main variables: HPV awareness and HPV vaccine awareness. Since too many variables can interfere with generating an unbiased dataset, identifying auxiliary variables related to demographic variables are not performed. Instead, identifying auxiliary variables related to main variables are performed. Identified auxiliary variables include internet use as a primary medical information source,<sup>19</sup> smoking,<sup>20</sup> psychological distress,<sup>21</sup> and cervical or breast cancer screening.<sup>22</sup>

After auxiliary variables are identified from the literature, researchers examine the relationships among each auxiliary

variable and each main research variable to finalize the auxiliary variables for the imputation. Correlation, t-test, or chi-square test can be used and variables showing significant results are included in the imputation. Using the HPV variables from the HINTS data, internet use as a primary medical information source and cervical or breast cancer screening showed significant results among auxiliary variables. Thus, three variables can be used in the imputation phase: internet use as a primary medical information source; cervical cancer screening; and breast cancer screening.

### *Imputation Phase*

In the imputation phase, the number of datasets with estimations of missing values has traditionally been three to five.<sup>12</sup> However, Kenward, Carpenter<sup>23</sup> have suggested 100–200 datasets. Enders<sup>12</sup> also suggested large number of datasets because as more datasets are created, the standard error decreases. Most recently<sup>24</sup> study suggests that the number of datasets to be created should be calculated based on the fraction of missing information and the coefficient of variation.

Estimations of missing values are random by using a statistical algorithm with auxiliary variables. The estimation algorithm is based on the study design, the variables, or the missing rate.<sup>12</sup> For example, if researchers could not assume either missing completely at random or missing at random, then an algorithm using Markov chain Monte Carlo or regression method could be used.<sup>16</sup> If there is interaction among categorical and continuous independent variables, then multiple imputation by chained equations might be more accurate.<sup>25</sup> After the algorithms are used as part of the imputation phase, then random estimates for missing data is included in each dataset, resulting in each dataset having different data values for the missing values. For this example, internet use as a primary medical information source, cervical cancer screening, and breast cancer screening were included in this phase. If we created 10 datasets, then 10 complete datasets containing independent, dependent, and three auxiliary variables would be created. Each of the ten datasets would use different values for replacing the missing values because for each dataset missing values were replaced randomly.

### *Analysis Phase*

In this second phase, researchers apply the same statistical analysis method with the newly created datasets from the imputation phase. If 10 sets of data were created in the imputation phase, then the analysis phase will include conducting 10 different analyses. As a result, this phase produces multiple statistical results from each dataset.

For example, if a researcher wants to analyze regression to examine the association between HPV awareness and HPV vaccine awareness, then 10 different regressions will be conducted using each complete dataset. Consequently,

10 different datasets of regression parameter estimates and standard error will be calculated.

### *Pooling Phase*

In the final pooling phase, all statistical results from the analysis phase are combined.<sup>12</sup> The pooling phase calculates the average of parameter estimates and merges the estimated variability of standard errors.<sup>11</sup> For example, in this phase, the 10 different parameter estimates and standard errors from the analysis phase are pooled and represented as one set of results.

### *Dealing With Screening Questions*

As previously mentioned, screening questions can produce missing data values. In HINTS data, some HPV variables have a screening question. In this section, the HPV vaccine recommendation variable is used as an example of the main variable to explain the situation caused by the screening question.

The HPV vaccine recommendation variable is screened by the question: “Including yourself, is anyone in your immediate family between the ages of 9 and 27 years old?” The possible answer is either “yes” or “no.” In total, 3,474 respondents answered the screening question, with .9% missing rate. The number of respondents who answered “yes” were 1,349 (38.5%), and “no” were 2,125 (60.6%). When respondents answered “yes” on the screening question, they were asked to answer the HPV vaccine recommendation question: “In the last 12 months, has a doctor or health care professional recommended that you or someone in your immediate family get an HPV shot or vaccine?” The possible answer is either “yes,” “no,” or “don’t know.” The HINTS was exclusively conducted by mail.<sup>10</sup> Therefore, it is possible that respondents could answer the HPV vaccine recommendation question even if they did not answer “yes” on the screening question. Indeed, among the respondents who did not answer “yes” (2,155, 61.5%), many of them (1,784, 50.9%) answered the HPV vaccine recommendation question. However, the HINTS reported these answers as missing data (coded –1, with the label of “inapplicable”) in the HPV vaccine recommendation variable. The missing data was 2,166 which is a missing rate of 61.8%.

If respondents answered that they have a family member aged 9–27 years old, then the possible answer for the main variable, HPV vaccine variable, could be “yes,” “no,” “don’t know,” or missing values. If someone answered “no” in the screening question, this is also valid, because the HPV vaccine recommendation question is systematically considered missing data. However, a problem occurs when the answer for the screening question is missing.

There is a gap in the research about situations that include screening questions with imputation. However, this paper suggests portioning out the cases that only answered “yes” in the screening question and use a portion of the dataset for

imputation and further analysis. In other words, in the example of HINTS data, only cases where respondents answered “yes” in the screening question, which is 1,349 (38.5%), would consider for imputation and further analysis.

There are three reasons supporting this suggestion. First, it is impossible to conduct multiple imputation twice. Suppose the researcher conducted multiple imputation first for the screening question, and then subsequently, conducted a second multiple imputation for the interested variable. As mentioned earlier, multiple imputation contains the process of statistical analysis. The first imputation will produce sets of statistical results such as parameter estimates, not a complete dataset. The results from the first multiple imputation cannot be used for another imputation method. Therefore, conducting multiple imputation twice is not possible, even if the concept seems plausible.

Second, nationally representative secondary data usually collects data based on stratified sampling, so the results need to incorporate sample weights.<sup>26</sup> If the first imputation is conducted with sample weight, the results will reflect the population, but they will not be based on the data sample itself. As described earlier, researchers have to apply sample weight twice as well as multiple imputation if there are missing values on the screening question. This is another example of why multiple imputation cannot be conducted twice.

Third, if the researcher conducted multiple imputation as a whole including screening questions and the target variable, then the computer program will generate estimates for the missing values on the interested variable without distinguishing whether missing data is caused by the screening question or a result of the survey recipients’ refusal to answer the question.

For example, when respondents did not answer the question asking if there is a family member who is from 9 to 27 years of age, the possible estimates of the missing values for the screening question are either yes or no. If the estimation is “yes,” then the missing value in HPV vaccine recommendation variable could be estimated to be “yes,” “no,” or “don’t know.” However, if the response to the question about the family member is “no,” then there is no logical answer for an estimate for the HPV vaccine recommendation variable. Even if the researcher assumes the proper answer of the HPV vaccine recommendation variable would be “no” or “don’t know” when the answer for the screening question is “no,” there is no way to prohibit the estimate of “no” on screening question and “yes” on the HPV vaccine recommendation variable. Therefore, if there are missing values for the screening question, it is recommended to extract cases that have passed the screening question, and use these cases for imputation and further analysis with sample weights.

## Discussion

This paper discussed the causes of missing data in secondary datasets and possible approaches for handling missing data

using multiple imputation. The three causes of missing data are (a) survey structure, (b) refusal to answer, and (c) insufficient knowledge about questions.

This paper explored the literature for recommendations of how to handle missing data when the cause is refusal to answer a question because of insufficient knowledge. When there is small portion of the missing values, less than 10%, then complete cases analysis can be used. However, because of bias and wastefulness, when the missing rate is greater than 10%, multiple imputation is commonly used. Multiple imputation consists of three phases: imputation phase, analysis phase, and pooling phase. Researchers need to select auxiliary variables prior to conducting the multiple imputation. Auxiliary variables are used for the first imputation phase. In the imputation phase, a certain number of datasets with estimates of missing values are created. The analysis phase applies the same statistical analysis method for each dataset created in the imputation phase. In final pooling phase, all statistical results from each analysis are pooled.

This paper also explored the issue of handling missing data caused by the survey structure. Since researchers performing secondary data analysis did not design the survey, missing data caused by the survey structure can occur. There is a gap in the literature related to handling missing data when the cause is the survey structure. This paper suggests portioning out the cases that answered “yes” in the screening question and using that portion of the dataset for analysis. The rationale for this suggestion is that it is impossible to conduct multiple imputation twice, and it is also impossible to prohibit logical errors during the estimation of missing values. In addition, nationally representative secondary data needs to incorporate sample weights; applying sample weight twice is also impossible.

## Limitations

This study has limitations. First, this study did not conduct a rigorous systematic literature review. This may limit findings of the other proper methods for handling missing data. Second, the example used in this paper described the multiple imputation method using existing secondary data, which adds a layer of complexity to the analysis. Solutions to missing data could be easier to understand by using more complex analysis methods. Third, this study focused on complete case analysis and multiple imputation and did not address other methods for handling missing data. Researchers could select the proper method based on the situation. Further review on other methods is suggested.

## Conclusion

This paper proposes possible reasons for missing data in a secondary data set, describes how to conduct multiple imputation, and addresses problems related to missing data at the level of a screening question. Many healthcare studies use

secondary data analysis. Because missing data affects the results of analysis, it is important to select appropriate methods for handling missing data. This manuscript would be helpful to conduct the future research encountering the missing data in the future secondary data analysis studies by explaining the steps of conducting multiple imputation.

Despite this decent method for handling missing data, problems could occur when there are missing data at the level of a screening question. There is a lack of evidence surrounding this issue, but the approach of analyzing only the portion of the dataset that includes answers to the screening question seems to be the best solution to date. By explaining the three phases of multiple imputation, the author's goals include supporting future method-focused research on improvements for addressing missing data in secondary data analysis.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iD

Soojung Jo  <https://orcid.org/0000-0002-2129-4373>

### References

- Vartanian TP. *Secondary Data Analysis*. New York: Oxford University Press; 2010.
- O'Connor S. *Secondary Data Analysis in Nursing Research: A Contemporary Discussion*, Vol 29. Los Angeles, CA: SAGE Publications Sage CA; 2020:279-284.
- Cheng HG, Phillips MR. Secondary analysis of existing data: Opportunities and implementation. *Shanghai archives of psychiatry*. 2014;26(6):371-375.
- Wickham RJ. Secondary Analysis Research. *Journal of the advanced practitioner in oncology*. 2019;10(4):395-400.
- Gloyd S, Wagenaar BH, Woelk GB, Kalibala S. Opportunities and challenges in conducting secondary analysis of HIV programmes using data from routine health information systems and personal health information. *J Int AIDS Soc*. 2016;19(5S4):20847.
- Berglund P. *An Introduction to Multiple Imputation of Complex Sample Data Using SAS® V9*. 2. Seattle, Washington: SAS global forum; 2010:2010.
- Shoemaker PJ, Eichholz M, Skewes EA. Item Nonresponse: Distinguishing between don't Know and Refuse. *Int J Publ Opin Res*. 2002;14(2):193-201.
- Kim HJ, Fredriksen-Goldsen KI. Nonresponse to a question on self-identified sexual orientation in a public health survey and its relationship to race and ethnicity. *American journal of public health*. 2013;103(1):67-69.
- Pridham G, Rockwood K, Rutenberg A. Strategies for handling missing data that improve Frailty Index estimation and predictive power: Lessons from the NHANES dataset. [published online February 01, 2022]. *GeroScience*. 2022.
- Health Information National Trends Survey. *About HINTS*. Published n.d. Accessed Dec 27, 2019 <https://hints.cancer.gov/about-hints/Default.aspx>.
- Allison PD. Missing data techniques for structural equation modeling. *J Abnorm Psychol*. 2003;112(4):545-557.
- Enders CK. *Applied Missing Data Analysis*. New York: Guilford press; 2010.
- Langkamp DL, Lehman A, Lemeshow S. Techniques for handling missing data in secondary analyses of large surveys. *Academic pediatrics*. 2010;10(3):205-210.
- Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. *BMC Med Res Methodol*. 2017;17(1):162.
- Zhang Z. Missing data imputation: focusing on single imputation. *Ann Transl Med*. 2016;4(1):9.
- Lee JH, Huber JC Jr. Evaluation of multiple imputation with large proportions of missing data: How much is too much? *Iran J Public Health*. 2021;50(7):1372-1380.
- Hardt J, Herke M, Leonhart R. Auxiliary variables in multiple imputation in regression with missing X: A warning against including too many in small sample research. *BMC Med Res Methodol*. 2012;12(1):184.
- Pedersen A, Mikkelsen E, Cronin-Fenton D, et al. Missing data and multiple imputation in clinical epidemiological research. *Clin Epidemiol*. 2017;9:157-166.
- Miller LMS, Bell RA. Online Health Information Seeking. *J Aging Health*. 2012;24(3):525-541.
- Wigfall LT, Sherman LD, Garney WR, Patterson MS, Ishino FAM, Vadaparampil ST. Are health care providers making the most of patient encounters to promote HPV vaccination among cigarette smokers? *Patient education and counseling*. 2020;103(1):180-188.
- Berry HG, Disckind BB, Reichard A, Ruiz S. Health characteristics and outcomes of caregivers in the United States: An analysis of the 2017 health information national trends survey (HINTS). *Disabil Health J*. 2020;13(1):100821.
- Kim H, Mahmood A, Carlton E, Goldsmith J, Chang C, Bhuyan S. Access to personal health records and screening for breast and cervical cancer among women with a family history of cancer. *J Cancer Educ*. 2020;35(6):1128-1134.
- Kenward MG, Carpenter J. Multiple imputation: Current perspectives. *Stat Methods Med Res*. 2007;16(3):199-218.
- Von Hippel PT. How many imputations do you need? A two-stage calculation using a quadratic rule. *Socio Methods Res*. 2020;49(3):699-718.
- Javadi S, Bahrapour A, Saber MM, Garrusi B, Baneshi MR. Evaluation of four multiple imputation methods for handling missing binary outcome data in the presence of an interaction between a dummy and a continuous variable. *Journal of Probability and Statistics*. 2021;2021:6668822.
- Stapleton LM. The incorporation of sample weights into multilevel structural equation models. *Struct Equ Model: A Multidiscip J*. 2002;9(4):475-502.