Differentiated Effects of Data Analyses on Between- and Within-Imputation Variances in Multiple Imputation[1]

Qiyuan Pan, Rong Wei, and Yulei He
National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, Maryland, USA
Contact email: qpan@cdc.gov

Abstract

In multiple imputation (MI), the total variance (T) is estimated by U+(1+1/m)B, where U is the within-imputation variance, B the between-imputation variance, and m the number of imputations. The expected value of U is not affected by a proper MI, whereas the extra variance B can be captured only by MI but not by single imputation (SI). Whether B is large enough to cause a meaningful change in T may have an effect on people's perspective towards the value of MI as compared to SI. This paper evaluates how data analysis affects the impact of MI (I_{MI}), measured as $I_{MI} = 100(B/T)^{1/2}$. MI trials were conducted using the data of the 2012 Physician Workflow Mail Survey. Difference in analytic models had differentiated effects on B and U. Our results suggest that, for the same MI and the same data, I_{MI} may be negligible (<1%) in one analysis but substantial (>5%) in another.

Keywords: Impact of multiple imputation; Multiple imputation; Missing data; Between-imputation variance; National Ambulatory Medical Care Survey.

1. Introduction

Multiple imputation (MI) has become the most popular approach in dealing with the missing data (Carpenter and Kenward 2013; Rezvan et al. 2015). However, not all data collectors and users are convinced that it is worthwhile to adopt MI due to inconsistencies in its effectiveness. Many have reported that MI effectively reduced bias (Dohoo 2015; Walani et al. 2015), but others found that MI had little effect (Pan and Shimizu 2009; White and Carlin 2010; Twisk et al. 2013). This paper studies how different analyses on the same MI and the same data may affect the impact of the MI on the final result.

Many researchers pointed out that MI must be compatible with analysis (Kontopantelis et al. 2017; Ludtke et al. 2017; Rawlings et al. 2017). This compatibility between MI and analysis, however, should not be interpreted as being that one can perform a unique analysis only for a particular MI. In fact, the complete datasets generated from the same MI may be legitimately analyzed in many different ways (Cattle et al. 2011; van Buuren 2012, section 2.3.4). This is particularly true for large national surveys such as National Ambulatory Medical Care Survey (NAMCS). Once the data of these surveys are released to the public, they are subject to many different analyses by various data users all over the world.

In general, the application of MI technique is made up of three steps as follows (van Buuren 2012, section 6.4):

---

[1] The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the National Center for Health Statistics, Centers for Disease Control. This work was performed under employment by the US federal government; the authors did not receive any outside funding.

Step 1. Imputation: Impute the missing data for m times, resulting in m complete datasets, where m is the number of imputations of the MI.
Step 2. Analysis: Analyze each of the m datasets.
Step 3. Pooling: Integrate the m results into the final result.

Step 1 attracts the most researches because people want to know how MI should be properly carried out for various data situations. To date little research has been published on the relationship between data analysis, i.e. step 2, and the impact of MI on the final results.

Rubin's rule for pooling the m results to form the final result is represented by the following three equations (Rubin 1987):

$$B = \frac{1}{m-1}\sum_1^m (Q_i - \overline{Q})^2 \quad (1)$$

$$U = \frac{1}{m}\sum_1^m U_i \quad (2)$$

$$T = U + \left(1 + \frac{1}{m}\right) B \quad (3)$$

where B, U, and T are the between-, within-, and total variances, respectively, Q is the quantity of interest, and the subscript *i* stands for the *i*th imputation of the MI. The expected value of U is not affected by a proper MI (Rubin 1987). The gain from MI as compared to the single imputation (SI) is that MI makes it possible to estimate B whereas it is impossible to estimate B in SI. Therefore the impact of MI can be measured by the size of B relative to U and T. In this paper, I_{MI}, the ratio of percentage B/T in standard error unit was used as the measure of the impact of MI:

$$I_{MI} = 100\sqrt{B/T} \quad (4)$$

For the same data and the same MI, the difference in I_{MI} would be possible only if step 2, the analysis of the data, has differentiated effects on B and U, which was what we found and are reporting in this paper. The results from this research may help data programs to decide whether it is worthwhile to adopt MI.

2. Methods

MI trials were carried out using the data of the 2012 Physician Workflow Mail Survey (PWS), a supplement of NAMCS. Descriptions of PWS and NAMCS are available in Jamoom et al. (2012), Lau at el. (2016) and in the NAMCS webpage https://www.cdc.gov/nchs/ahcd/index.htm. The study had four treatment factors, namely the imputation variable (ImpV), the missing data percentage (δ), the imputation covariant variable (ImpCoV), and the analytic variable (AnaV). ImpV is the variables whose missing values were imputed. ImpV had three values, SIZE5, SIZE20 and SIZE100. They were the variables of the physician's practice size in different scales. The values of SIZE100 ranged from 1 to 100. SIZE5 was derived from SIZE100 by recoding the values into five levels, and SIZE20 was derived from SIZE100 by top-coding the > 20 values into 20.

Treatment factor δ had two values, 4% and 29%. The 2012 PWS had 29% missing data due to item nonresponse for SIZE100. The 4% missing level was obtained after the missing values were replaced with the nonmissing values available in the 2011 PWS for the same physician. SIZE5 and SIZE20 had the same missing data profile as SIZE 100. Treatment factor ImpCoV refers to the variables that were used as covariates in imputation, and treatment. Both ImpCoV and AnaV had the following four values: CK (the control), REGION (RG) (region where the physician's office was located), PRIMEMM (EM) (the physician's primary employment type), and DERIVED (DR) (a 3-value derived variable that was highly correlated with the ImpV). For ImpCoV, CK was the MI without a covariate, and for AnaV, CK was the analysis without an analytic variable. The quantity of interest (Q) was the mean of the

physician's practice size for the nation or for a specified sector (e.g. REGION 1) as represented by the scale of SIZE5, SIZE20, or SIZE100.

Hot deck imputation (Siddique and Belin 2008; Andridge and Little 2010) was used for the MI trials. The nonmissing portion of the δ = 4% datasets were used as the donors for imputing both δ = 4% and 29% missing values in the imputation. The m values chosen for the MI trials was 60. Twenty MI replicates were conducted for each treatment combination. Unweighted data were used in analyses. All estimates were for research purpose only.

3. Results and discussion

3.1 Differentiated effects of analyses on B and U

For the same data and MI, the square root value of B or U from the analysis with no AnaV (i.e. AnaV = CK) are compared to the mean of the square root values of B or U, i.e. $B^{1/2}$ and $U^{1/2}$, from the analyses with AnaV = RG, DR, or EM. The results for ImpV = SIZE5 were presented in Figure 1, and those for ImpV = SIZE20 were presented in Figure 2. Each of Figures 1 and 2 has four graphs, which were labeled as aB, aU, bB, and bU. Graphs aB and aU were the $B^{1/2}$ and $U^{1/2}$ data, respectively, for δ = 4%, and graphs bB and bU were the $B^{1/2}$ and $U^{1/2}$ data, respectively, for δ = 29%.

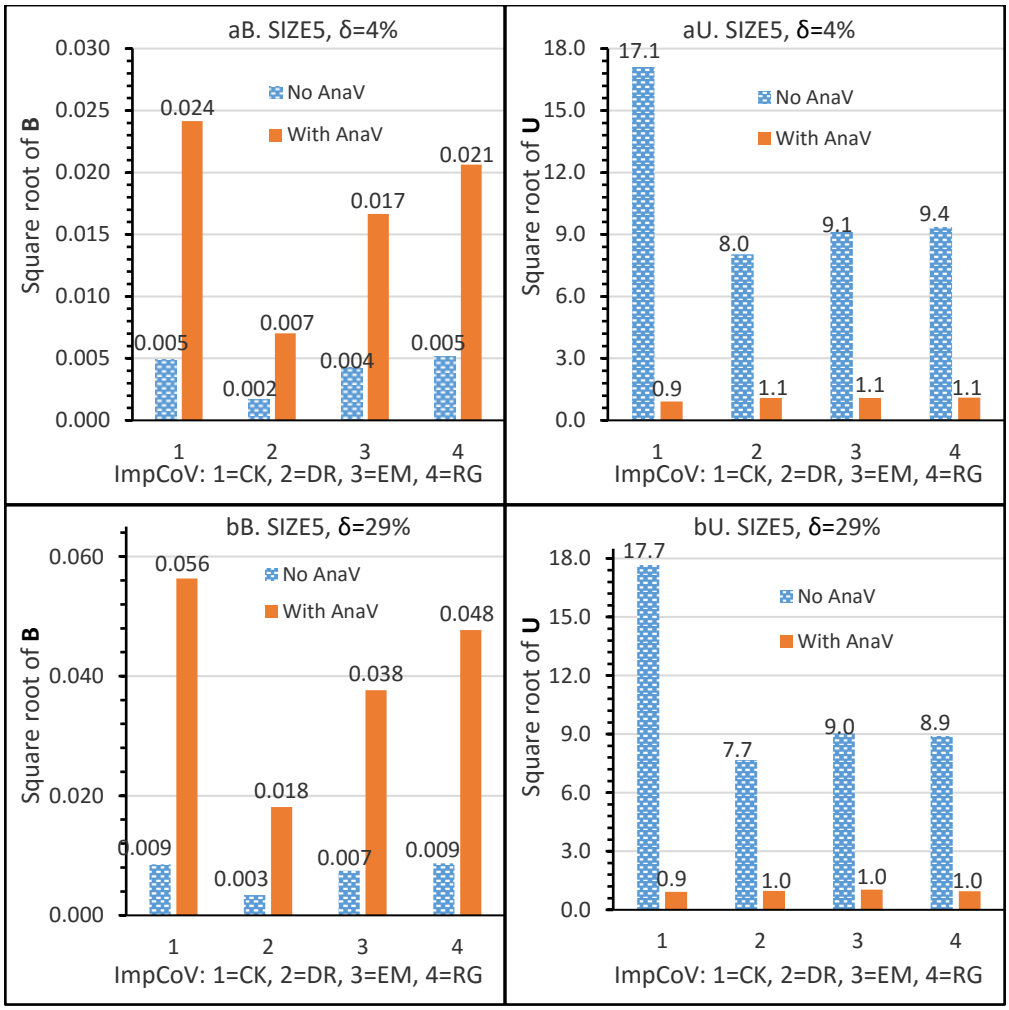

Figure 1. Effect of AnaV versus no AnaV in analysis on square root of B and square root of U for SIZE5

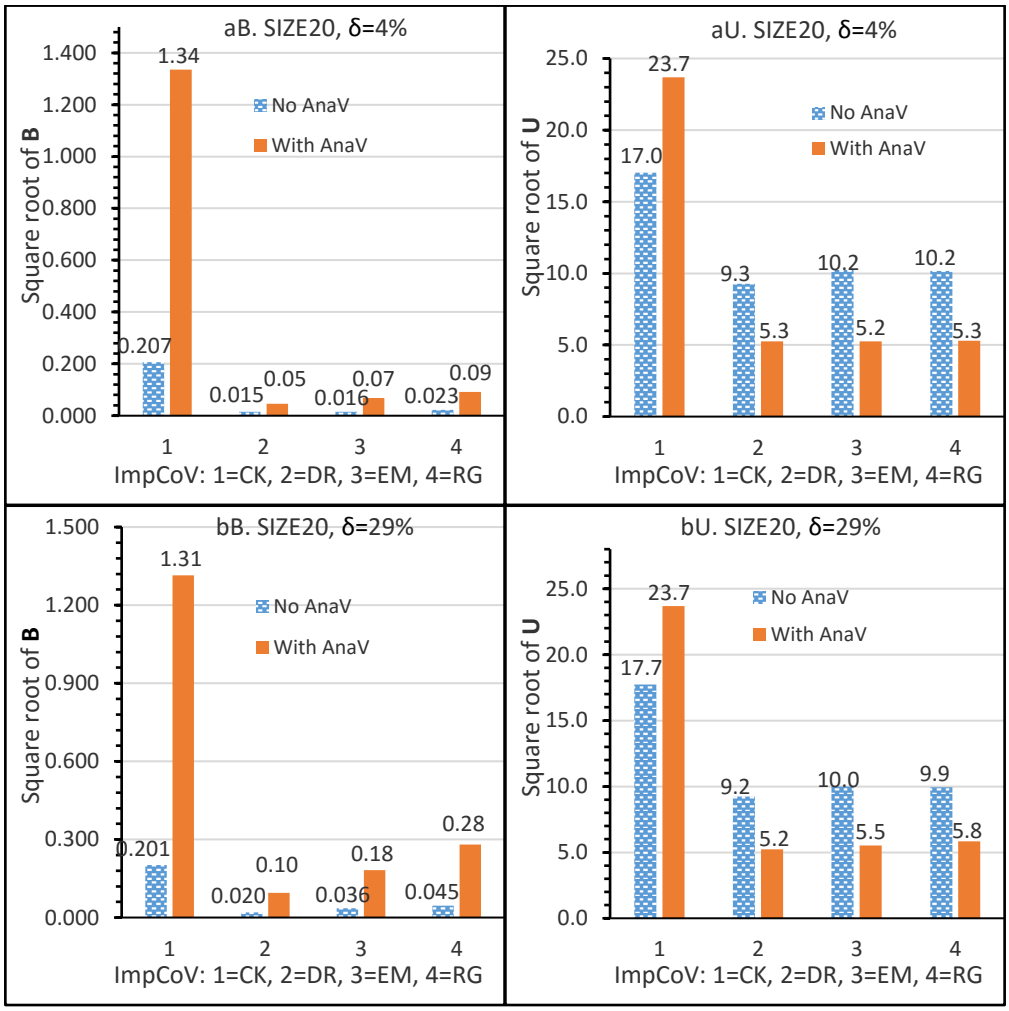

Figure 2. Effect of AnaV versus no AnaV in analysis on square root of B and square root of U for SIZE20

How the data were analyzed discriminately affected B and U. Take graphs aB and aU of Figure 1 for example. $B^{1/2}$ value for the "with AnaV" treatment was 3 to 5 times higher than that for the "no AnaV" treatment. But for the same data (ImpV = SIZE5, δ = 4%) and the same MI (ImpCoV = CK, DR, EM, or RG), $U^{1/2}$ for the "with AnaV" treatment was just one seventh to one nineteenth of that for the "no AnaV" treatment (Figure 1 aB and aU). The $B^{1/2}$ and $U^{1/2}$ change patterns were similar for ImpV = SIZE5 and δ = 29% (Figure 1 bB and bU).

The $B^{1/2}$ and $U^{1/2}$ change patterns of SIZE20 were somewhat different from those of SIZE5. For ImpCoV = CK, both $B^{1/2}$ and $U^{1/2}$ increased as the analytic model shifted from "no AnaV" to "with AnaV" (Figure 2), which was quite different from SIZE5 (Figure 1). Even though both $B^{1/2}$ and $U^{1/2}$ were higher for "with AnaV" than for "no AnaV" for ImpCoV = CK, the percentage increase was much higher for $B^{1/2}$ than for $U^{1/2}$ (Figure 2). Therefore, the statement that analysis differentially affects B and U still holds true for SIZE20 for ImpCoV = CK. For rest ImpCoV values, i.e. ImpCoV = DR, EM, or RG, $B^{1/2}$ increased but $U^{1/2}$ decreased as the analytic model shifted from "no AnaV" to "with AnaV" (Figure 2), a change pattern similar to that of SIZE5 (Figure 1). Data of SIZE100 were not shown to avoid redundancy because the message from SIZE100 data was very similar, as being indirectly suggested by the I_{MI} data in graphs a and b of Figure 3.

3.2 Effect of analysis on I_{MI}

The I_{MI} data for SIZE100 and SIZE5 were presented in Figure 3. Each of the four graphs in Figure 3 represents one particular data set constituted by a specific ImpV-δ combination. Four different MIs as defined by the four ImpCoV values were included in each graph. All together there are 16 data-MI combinations whose data of the I_{MI} comparison between the two types of analyses , "no AnaV" vs. "with AnaV", are presented in Figure 3. For the same data and MI, I_{MI} was always substantially greater for "with AnaV" than for "no AnaV" for all 16 data-MI combinations (Figure 3). The I_{MI} data of SIZE20 are not presented to avoid redundancy. As indirectly suggested by the $B^{1/2}$ and $U^{1/2}$ data in Figure 2, the I_{MI} change pattern for SIZE20 is very similar to that for SIZE5 and SIZE100.

The standard error (SE) for making the statistical inferences (e.g. constructing a confidence interval) is the square root of T (Rubin 1987). I_{MI} can be interpreted as the percentage change of SE caused by B. Let's say that a <1% I_{MI} value as indicating an ignorable impact of the MI, a 1%≤ I_{MI}≤ 5% value as indicating a borderline impact of the MI, and a ≥5% I_{MI} value as indicating a substantive impact of the MI. In two of the 16 data-MI situations in Figure 3, i.e. ImpCoV = CK and RG for SIZE5 and δ = 29%, the impact of MI was ignorable in one analysis and substantive in another for the same data and MI (Figure 3 d). In one data-MI situations, i.e. ImpCoV = DR for SIZE5 and δ = 4%, the impact of MI was ignorable for both analyses (Figure 3 c). In the remaining 13 data-MI situations, the impact of

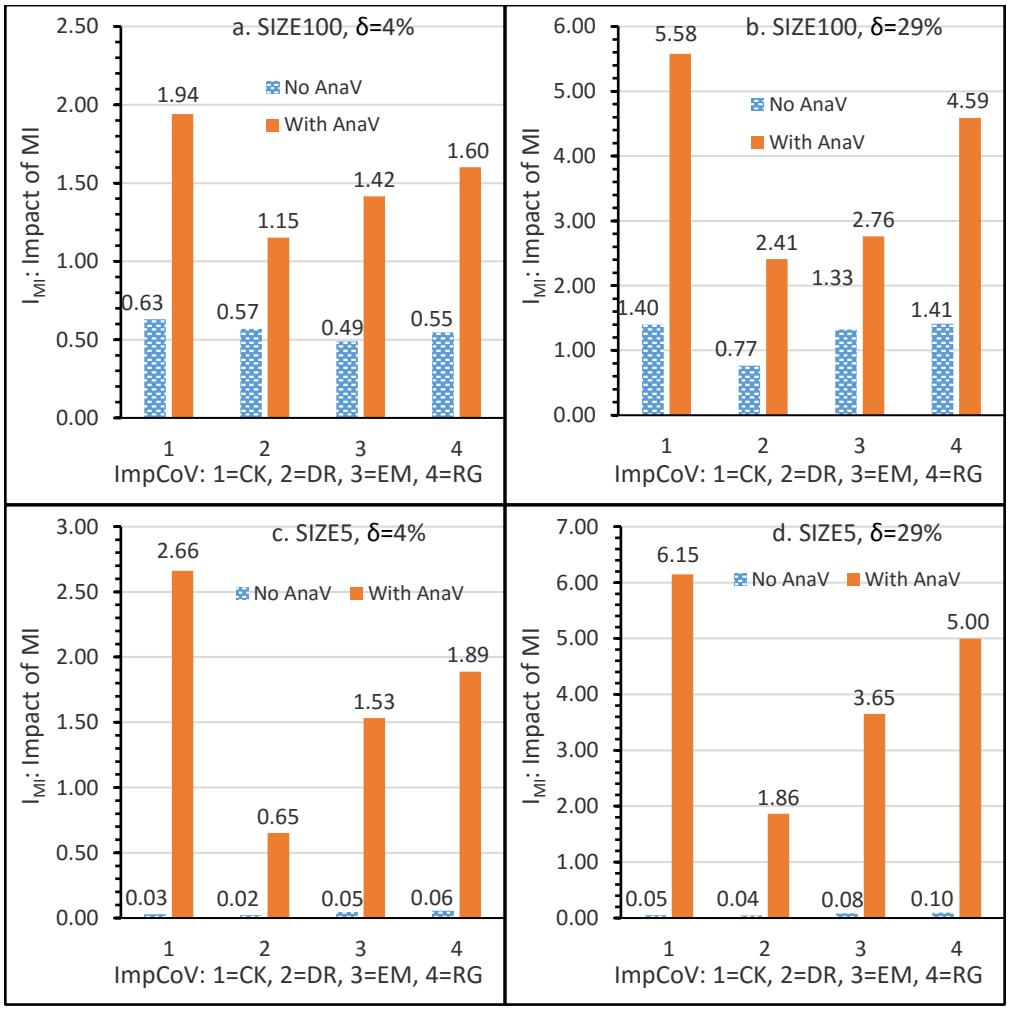

Figure 3. I_{MI} comparison between the analysis with AnaV and that without AnaV

MI was ignorable in one analysis and not ignorable in another for the same data and the same MI (Figure 3).

4. Conclusions

The results of MI trials performed on PWS12 data suggest that data analysis can discriminately affect B and U. For the same data and the same MI, I_{MI} can be very different in different analyses. The impact of MI may be too small to cause any meaningful change in the statistical conclusions in one analysis but can be large enough to cause substantive changes in the conclusions in another analysis. Therefore, where different analyses are possible by different data users, one cannot conclude that the MI would not be worthwhile just because of a small I_{MI} in his or her analyses.

References

Andridge, R. R., and Little, R. J. (2010), "A Review of Hot Deck Imputation for Survey Non-response," International Statistical Review, 78, No. 1, 40–64.

Carpenter, J. and Kenward, M. (2013), Multiple Imputation and its Application, John Wiley & Sons.

Cattle, B. A., Baxter, P. D., Greenwood, D. C., Gale, C. P., West, R. M. (2011), "Multiple imputation for completion of a national clinical audit dataset," *Statistics in Medicine*, 30(22): 2736-2753.

Dohoo, I. R. (2015), "Dealing With Deficient and Missing Data," Preventive Veterinary Medicine, 122 (1-2), 221-228.

Jamoom, E., Beatty, P., Bercovitz, A., Woodwell, D., Palso, K., and Rechtsteiner, E. (2012), "Physician adoption of electronic health record systems: United States, 2011", *NCHS data brief*, No. 98, Hyattsville, Maryland, USA, National Center for Health Statistics.

Kontopantelis, E., Ian R. White, Matthew Sperrin, and Iain Buchan (2017), "Outcome-sensitive multiple imputation: a simulation study," *BMC Medical Research Methodology*, 17(1): 2.

Lau, D.T., McCaig, L.F., and Hing, E. (2016), "Toward a More Complete Picture of Outpatient, Office-Based Health Care in the U.S.: Expansion of NAMCS", *Am. J. Prev. Med.*, 51(3):403-409. doi: 10.1016/j.amepre.2016.02.028.

Ludtke, O., Robitzsch, A., and Grund, S. (2017), "Multiple imputation of missing data in multilevel designs: A comparison of different strategies," *Psychological Methods*, 22(1): 141-165.

Pan, Q., and Shimizu, I. (2009), "Imputation Variance Estimation by Multiple Imputation Method for the National Hospital Discharge Survey," In *JSM Proceedings*, Alexandria, VA: American Statistical Association. pp. 1106-1114.

Rawlings, A. M., Sang, Y., Sharrett, A. R., Coresh, J., Griswold, M., Kucharska-Newton, A. M., Palta, P., Wruck, L. M., Gross, A. L., Deal, J. A., Power, M. C., and Bandeen-Roche, K. J. (2017), "Multiple imputation of cognitive performance as a repeatedly measured outcome," *European Journal of Epidemiology*, 32(1): 55-66.

Rezvan, P. H., Lee, K. J., and Simpson, J. A. (2015), "The rise of multiple imputation: a review of the reporting and implementation of the method in medical research," *BMC Medical Research Methodology*, 15: 30.

Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley & Sons, pp. 1-23 and pp. 75-147.

Siddique, J. and Belin, T. R. (2008), "Multiple imputation using an iterative hot-deck with distance-based donor selection," *Statistics in Medicine*, 27(1): 83-102.

Twisk, J., de Boer, M., de Vente, W., and Heymans, M. (2013), "Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis," *Journal of Clinical Epidemiology*, 66(9): 1022-1028.

Van Buuren, S. (2012), "Chapter 2 - Multiple imputation," *Flexible Imputation of Missing Data*, Chapman and Hall / CRC Press, Boca Raton, pp. 25-52.

Walani, S. R., and Cleland, C. M. (2015), "The multiple imputation method: a case study involving secondary data analysis," *Nurse Researcher*, 22(5): 13-19.

White, I. R., and Carlin, J. B. (2010), "Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values," *Statistics in Medicine* 29(28): 2920-2931.