Comparison of Alternative Imputation Methods in the National Teacher and Principal Survey [1]

Sarah Dial, Jacob Enriquez, Bonnie Moore, Svetlana Mosina,
US Census Bureau, 4700 Silver Hill Road, Washington DC 20233

Abstract
The Schools and Staffing Survey (SASS) is undergoing a redesign and will be called the National Teacher and Principal Survey (NTPS) during its next administration, to be conducted during the 2015-16 school year. As part of this redesign, it is of interest to determine if multiple imputation methods or administrative records could replace or supplement the current hot deck imputation procedure. One research objective is to determine if the coverage and quality of the administrative records is sufficient to use as an alternative imputation source. The other objective is to explore additional suitable imputation techniques with and without using administrative records. This paper will discuss the direct assignment imputation method, in which administrative records will be used to assign values to missing data for the matching SASS record. Other imputation methods that will be explored include predictive mean matching imputation and propensity score imputation. To evaluate the proposed imputation methods we will look at several metrics that examine the bias of the predictive, distributional, and estimation accuracy of each of the methods.

Key Words: Imputation, Administrative Records, Item Nonresponse

1. Introduction

Nonresponse is a common problem in surveys that can lead to a loss of precision and bias in estimates. Item nonresponse, in particular, occurs when partial data is collected for a respondent with some items, or questions, missing. Data entry errors can also lead to item nonresponse. One solution for item nonresponse is replacing missing data with data from administrative records. Another solution for item nonresponse is imputation. Imputation involves substituting missing values with a single value through the process of single imputation or substituting missing values with a set of plausible values through the process of multiple imputation. Single imputation methods do not account for the variability in each imputation because these methods disregard the fact that each imputed value is selected from a pool of plausible replacements for each missing value, as opposed to the single true known value. As a result, such analysis of imputed data may yield underestimated standard errors. These methods could lead to making incorrect inferences on the statistical significance of coefficients (Schenker, 2006). Multiple imputation methods enable analysts to account for the variability due to the uncertainty caused by imputing values multiple or repeated times (Rubin, 1988).

[1] Disclaimer: This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed are those of the authors and not necessarily those of the U. S. Census Bureau.

The primary goal of this research is to determine if the 2011-12 SASS items listed in Table 1 below can potentially be replaced by administrative records data. These twenty-eight items were chosen because the response information is also available from an administrative records data source. The coverage and quality of the administrative data for these items are assessed using evaluation measures mentioned in a subsequent section.

Table 1: 2011-12 SASS School Questionnaire Items

Item Description (Number of Items in This Category)	Type of Question	Response Rate Percentage	Number of Respondents Eligible to Answer
Grades Offered (15)	Binary	99.97	7481
Total Enrollment (1)	Discrete	100.00	7481
Enrollment by Race (8)	Discrete	94.96*	7481
School Type (1)	Categorical	100.00	7481
Number of Full Time, Part Time, and Total Teachers (3)	Discrete	99.72*	7481

*Average response rate

The secondary goal of this research is to determine if the current imputation method for the items listed in Table 2 can be improved using alternative multiple imputation methods, both with and without the use of administrative records data. It is important to note the difference in the use of administrative records for items in Table 1 and Table 2. Administrative records are used for imputation for items in Table 2 and the remaining missing values will be imputed using the alternative multiple imputation methods, whereas administrative records are used to completely replace the items in Table 1. Markov Chain Monte Carlo (MCMC), regression, propensity score, and predictive mean matching imputation are the alternative imputation methods that are explored for items in Table 2. A mixture of items that have lower item response rates as well as those that have higher item response rates were chosen to determine the effectiveness of imputation methods on different levels of item nonresponse. The sample size was also considered when selecting these items to ensure that there would be enough cases for model fitting. The current single hot deck imputation method is compared to the alternative multiple imputation methods using evaluation measures mentioned in a subsequent section. The items in Table 2 are also referred to as response variables later in this paper.

Table 2: 2011-12 SASS School and Teacher Questionnaires Items

Item Description	Type of Question	Response Rate Percentage	Number of Respondents Eligible to Answer
Black Enrollment (without admin. data)	Discrete	93.84	7481
Black Enrollment (with admin. data)	Discrete	99.56	7481
Newly Hired Teachers	Discrete	96.85	7481
Pension Check (how much)	Continuous	72.51	905

2. Background

The purpose of the SASS has historically been to collect the information necessary to form a complete picture of American elementary and secondary education. Currently, SASS is conducted by the U.S. Census Bureau for National Center for Education Statistics (NCES) every four years to collect information on what is happening in K-12 public and private schools from both the administrator and teacher perspective. Information collected includes teacher demand, teacher and principal characteristics, and general conditions in schools. The redesigned SASS, which will be called NTPS, will have a different structure and sample than previous administrations of SASS; however, it will maintain the same focus on schools and their teachers and administrators that was traditionally held by the SASS (National Center for Education Statistics, 2013).

The 2011-12 SASS public school frame was built from the 2009-10 Common Core of Data (CCD) administrative data file. The CCD is a universe of public schools collected from state education agencies by NCES. The universe was modified to fit SASS definitions of a school by adding, deleting, and collapsing schools. For the 2011-12 administration of SASS, 11,000 public schools were sampled.

The current imputation method for the 2011-12 SASS is the single imputation hot deck donor method. The hot deck method matches a record containing missing data to a donor record within the same file. A match is made using a set of variables (called "matching variables") whose values are identical on both the imputed and the donor records. The missing data is then replaced with the data found on the donor record. Matching variables are recoded survey items or variables from the frame, and cannot have missing values. This method is based on the assumption that records with the same values for the matching variables will respond similarly to certain survey items. If more than one donor is found, one is selected at random. A record can be used as a donor a maximum of five times, unless there is a small pool of potential donors. Matching variables have been used for several administrations of SASS. After each missing item is imputed, the file goes through consistency edits to ensure that each imputed value is consistent with other survey responses.

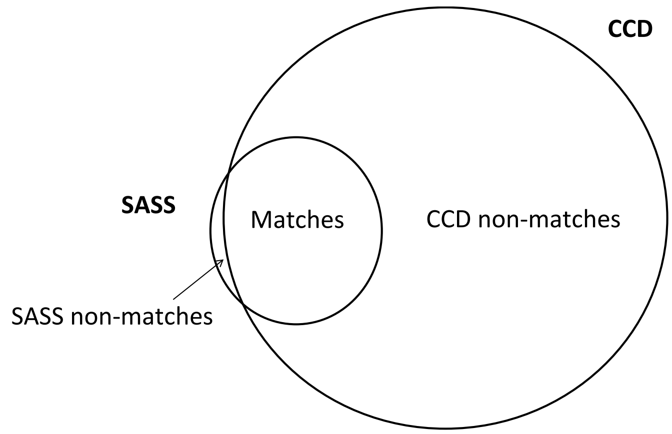
The concept of using donors has it drawbacks which includes not having enough donors. This occurs whenever no record meets the conditions that potential donors must satisfy, such as having the right values for the matching variables, selected as a donor no more than five times, and conditions specific to each survey question. Another problem of using donors is selecting a “bad” donor. Even with the use of matching variables, many times the donor selected is not similar to the record to be imputed. This can result in imputing outlier values. The only solution to these issues is to perform manual fixes, which are quite tedious and time consuming. This study also explores the use of imputation methods that do not require the use of donors such as MCMC and regression.

3. Methodology

3.1 Administrative Records Coverage and Quality

The administrative records data source that could potentially be used to replace SASS data items is the CCD. Hence, the quality and coverage of the 2009-10 CCD is evaluated in this research to determine if it could have been used to replace the 2011-12 SASS items. To determine the coverage of the CCD, the school-level matching rate and the rate of reported values on the CCD are calculated. The school-level matching rate is calculated by matching the 2011-12 SASS public school dataset to the 2009-10 CCD by a unique school identifier variable. Figure 1 below depicts the school-level matching between the SASS dataset and the CCD.

Figure 1: Graphical Depiction of the Matching Outcomes



The records that are in both the SASS and CCD datasets are the matches, or matching schools, as shown in Figure 1. Records that are on the SASS file, but are not on the CCD are SASS nonmatches. This quantity represents the undercoverage of the CCD. Records that are on the CCD, but are not on the SASS file are CCD nonmatches. This quantity represents the overcoverage of the CCD, which are not investigated further. Because the CCD is a universe of 103,959 schools and the SASS dataset is a sample of 11,000 schools, a large amount of CCD nonmatches is expected.

The rate of reported values is calculated for the matching schools between the SASS and CCD datasets by the following formula:

$$rate\ of\ reported\ values = \frac{number\ of\ schools\ on\ the\ CCD\ with\ a\ nonmissing\ value}{total\ number\ of\ matching\ schools}$$

If the rate of reported values is low, the administrative data will not be useful for replacement of the SASS data items in Table 1 because the item will have a high rate of missingness as a result.

To determine the quality of the administrative data, the relative difference of values is calculated and a paired t-test is performed. The relative difference of values is calculated for matching schools with a reported value on the CCD. The relative difference between the value on the SASS dataset and the value on the CCD dataset are calculated using the following formula:

$$relative\ difference = \frac{|SASS\ value - CCD\ value|}{SASS\ value}$$

The relative differences are categorized into exact matches (relative difference equals 0), matches with a relative difference of 5% or less, matches with a relative difference between 5% and 10%, and matches with a relative difference greater than 10%.

The paired t-test computes the difference between each of the paired CCD and SASS values and determines whether the mean of the differences is significantly different from zero using an alpha of 0.05. An item is considered a candidate for replacement with CCD values if the rate of reported values is high and the paired t-test shows no significant difference for the paired CCD and SASS values.

3.2 Truth Deck Dataset

Before creating the truth deck dataset, the models for each response variable in Table 2 are fitted. The items with the highest correlations to the response variable compared to the other potential items are put into the model. Then, the model is fitted for each response variable using the GLM procedure with the stepwise selection option in SAS®. These models are used as input for a simulation study. The simulation study involves creating an initial dataset of only completed cases, which will be referred to as the truth dataset. A random pattern of missingness is imposed on the response variable of each truth dataset at the same rate of missingness as that response variable (100 minus the response rate percentage listed in Table 2) to create the truth deck dataset. Three of the alternative imputation methods require a monotone missing data pattern. A monotone missing data pattern is achieved by executing MCMC for the covariates included in each model. MCMC is described in more detail in a later section. The generated missing values of the response variable are imputed five times using the MI procedure in SAS® for each of the alternative imputation methods. These methods are discussed later. This process is iterated 250 times.

The true values of the records are compared to the imputed values from the simulation. Then, the evaluation measures are produced, and one imputation method is chosen for imputation of the real missing values.

The missing at random (MAR) mechanism is assumed as it is defined by Rubin (1976) and Little and Rubin (1987): missing data values carry no information about probabilities of missingness and depends on the observed values of the 2011-12 SASS data.

3.3 Proposed Imputation Methodology

3.3.1 Direct Assignment Using Administrative Records

The direct assignment using administrative records imputation method first matches the SASS data to the CCD data by the unique school identifier. If the value on the CCD dataset is non-missing it will be assigned to the same response variable on the SASS dataset if it is missing. Other alternative imputation methods such as MCMC, regression, propensity score, and predictive mean matching can be used after the implementation of the direct assignment using administrative records imputation for the cases where the response variable is still missing.

3.3.2 Alternative Imputation Methods

A summary of the four alternative imputation methods is presented in Table 3. Detailed discussion of these methods can be found in the provided references.

Table 3. Summary of Alternative Imputation Methods.

Imputation Method (Reference)	**Description**
MCMC (Schafer, 1997)	Arbitrary missing pattern, generates pseudorandom draws from probability distributions via Markov chains, imputes with model-produced values.
Regression (Durrant, 2005; Hippel, 2012)	Monotone missing pattern, fitting a model that relates the response variable to the covariates, imputes with model-produced values.
Propensity Score (Rosenbaum and Rubin, 1983; Rubin, 1987)	Monotone missing pattern, conditional probability to assign value to imputed item using logistic regression, model-produced values.
PMM (Grannell and Murphy, 2011; Little, 1988)	Monotone missing pattern, linear prediction as a distance measure for the set of nearest neighbors (donors) consisting of the complete values, the respondent with the smallest distance metric is chosen as the donor, imputes with true values.

The MCMC, regression, propensity score and predictive mean matching imputation methods are all performed using the MI procedure in SAS®.

3.4 Evaluation Measures for Choosing Best Alternative Imputation Method

The following seven formulas were output from the truth deck dataset simulation and are used to assess the effectiveness of the four alternative imputation methods. The variable $\hat{Y}_i$ represents the imputed value and Y_i^{true} represents the observed, true value (Ziegelmeyer, 2011). The preferred method is the one that most often results in the lowest absolute values for the measures below.

The mean relative deviation is a predictive accuracy measure that quantifies how relatively close the imputed value is to the true value.

$$\boldsymbol{Mean\ Relative\ Deviation\ (MRD)} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|\hat{Y}_i - Y_i^{true}\right|}{Y_i^{true}}$$

The distributional accuracy measures below show how close the distribution of the imputed values is to the distribution of the true values. The percent of iterations with significant (alpha = 0.05) difference between alternatively imputed and observed means (% t-test sig.) is calculated in addition to the other distributional accuracy evaluation measures.

$$\boldsymbol{Quarter\ 1\ Bias\ (Q1\ Bias)} = (\hat{Y}_i)^{Q1} - (Y_i^{true})^{Q1}$$

$$\boldsymbol{Quarter\ 3\ Bias\ (Q3\ Bias)} = (\hat{Y}_i)^{Q3} - (Y_i^{true})^{Q3}$$

$$\boldsymbol{Median\ Bias\ (Med.Bias)} = (\hat{Y}_i)^{med} - (Y_i^{true})^{med}$$

$$\boldsymbol{Relative\ Bias\ (Rel.Bias)} = \frac{\frac{1}{n}\sum_{i=1}^{n} \hat{Y}_i \ - \ \frac{1}{n}\sum_{i=1}^{n} Y_i^{true}}{\frac{1}{n}\sum_{i=1}^{n} Y_i^{true}}$$

The following estimation accuracy measures show how well the imputed values reproduce the first and second moments of the distribution of the true values.

$$\boldsymbol{Mean\ Bias} = \frac{1}{n}\sum_{i=1}^{n} \hat{Y}_i \ - \ \frac{1}{n}\sum_{i=1}^{n} Y_i^{true}$$

$$\boldsymbol{Standard\ Deviation\ Bias\ (Std.Dev.Bias)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_i - \overline{\hat{Y}}\right)^2} - \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i^{true} - \overline{Y}^{true}\right)^2}$$

3.5 Evaluation Measures for Comparing Alternative Imputation Method to Current Hot Deck Method

Once an alternative imputation method is chosen and conducted for the missing values, the correlation structure and the t-test for means are used to compare the alternative imputation method to the current hot deck method. For the correlation structure, the data is split up into three groups: observed, imputed by alternative imputation method, and imputed by current hot deck method. The correlations between the covariates and response variable are computed for all three groups of data. The preferred imputation method is the one that preserves the correlation structure, that is, the correlations between the covariates and response variable of the observed data should be similar to the correlations after imputation.

A t-test for means is used to compare the means of the alternatively-imputed data and the observed data. Another t-test for means is used to compare the means of the hot-deck-imputed data and the observed data. This test determines whether the difference of the means of the two groups of data are significantly different from zero. An alpha of 0.05 is used. The preferred imputation method is the one whose mean is not significantly different than the mean of the observed data, thereby preserving the mean of the data after imputation.

4. Findings

4.1 Administrative Records Coverage and Quality

Out of 11,000 schools on the SASS dataset, 10,614 schools were on the 2010-11 CCD, which is a 96.5% school-level matching rate. This means that the school-level coverage of the CCD is very high. This is not unexpected since the SASS public school sample is drawn from the CCD universe. A 100% match is not achieved because schools are added, deleted, and collapsed on the CCD universe before SASS sample selection.

Table 4 below shows the results for the nine items with the highest percentage of reported values on the CCD. The remaining nineteen items that are listed in Table 1 have percentages of reported values on the CCD less than 42%, which would leave these items with a high rate of missingness. Consequently, those nineteen items will not be considered for replacement by administrative records data. Although the School Type item has no missing values on the CCD, it will not be considered for replacement either because the definition of school types on the CCD does not match the definitions of school types on the SASS.

Table 4: Rate of Reported Values on the CCD

Item Description	Percentage of Reported Values on CCD
School Type	100.00
Total Enrollment	98.41
Enrollment by Race	
Hispanic Enrollment	98.31
White Enrollment	98.31
Black Enrollment	98.31
Asian Enrollment	98.31
American Indian/Alaskan Native Enrollment	98.31
Total Race Enrollment	98.31
Total Teachers	97.84

Figure 2 and Table 5 show the results of the assessment of the quality of the administrative data. Figure 2 displays the percentage of relative differences between the SASS and CCD Values for the items listed in Table 4, excluding School Type. The abbreviation AI/AN in Figure 2 is for the American Indian/Alaskan Native Enrollment item. Ideally, there would be a low percentage of relative differences that were greater than 10%. Unfortunately, Figure 2 shows that every variable except for Total Enrollment and Total Race Enrollment have a large proportion of relative differences that are greater than 10%.

Figure 2: Percentage of Relative Differences between SASS and CCD Values

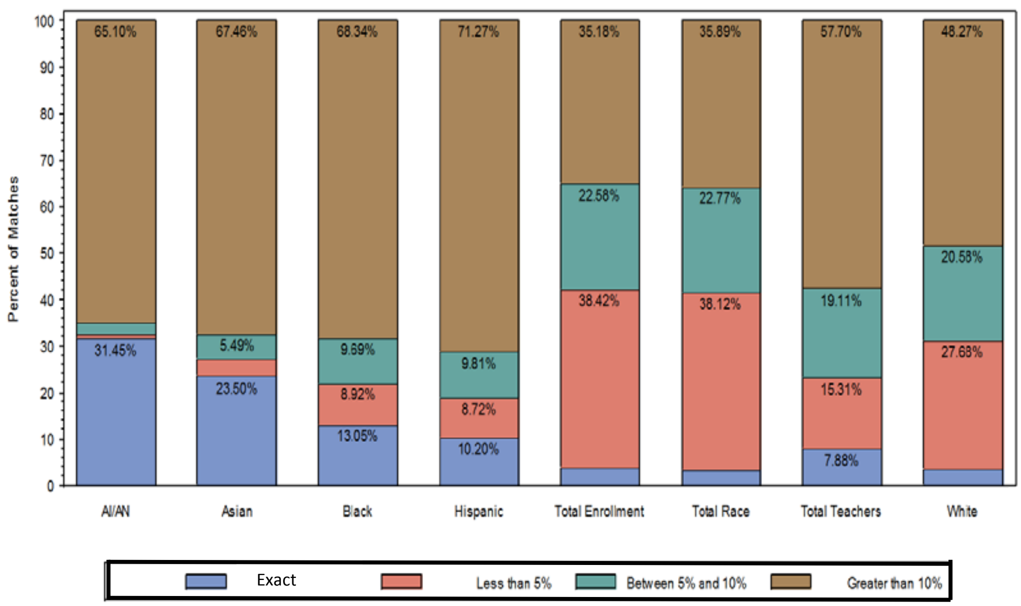


Table 5 below shows the p-values of the paired t-test of the CCD and SASS values. The average difference between the SASS Black Enrollment and the CCD Black Enrollment is only 0.37, which is not significantly different from zero as indicated by the t-test using an alpha of 0.05. The other enrollment by race items show a significant paired difference between SASS values and CCD values. Note that the average difference between SASS and CCD is also low for the American Indian/Alaskan Native Enrollment at -0.73; however, the standard deviation is much smaller. Consequently, the t-test is showing significant difference.

Table 5: Paired t-test Results

Item Description	Mean Difference	Std. Dev.	N	Pr > |t|
Total Enrollment	-24.85	221.70	7109	< .0001
Enrollment by Race				
Hispanic Enrollment	-9.48	93.56	6676	< .0001
White Enrollment	-8.37	147.90	6662	< .0001
Black Enrollment	-0.37	52.22	6674	.5671
Asian Enrollment	1.58	29.83	6682	< .0001
American Indian / Alaskan Native Enrollment	-0.73	14.94	6696	< .0001
Total race Enrollment	-29.89	221.50	7102	< .0001
Total Teachers	-4.45	13.46	7068	< .0001

4.3 Statistical Models

Regression models are developed to identify covariates for the imputation process. Table 6 shows the models that are fitted for items listed in Table 2 using the GLM procedure in SAS® with the stepwise selection option. The adjusted R-squared is also shown in Table 6. The adjusted R-squared values for the two Black Enrollment models are satisfactory at 0.89 and higher. The Newly Hired Teachers and Pension Check items, on the other hand, had low adjusted R-squared values at 0.27 and lower. Although a low adjusted R-squared generally means that the model may not provide a good fit, it was the best adjusted R-squared that could be achieved.

Table 6: Statistical Models Used

Item Description	Covariates in Model	Adj. R^2
Black Enrollment (without admin. data)	CCD Black Enrollment, Total Teachers, CCD Free and Reduced Lunch, Number of Vice Principals, Number of Black Teachers	.8905
Black Enrollment (with admin. data)	CCD Black Enrollment, Total Teachers, CCD Free and Reduced Lunch, Number of Vice Principals, Number of Black Teachers	.8933
Newly Hired Teachers	White Enrollment, Black Enrollment, Total Teachers, Number of Vice Principals, Number of Custodial and Security, Number of Students with IEP because of Special Needs	.2619
Pension Check	Highest Degree Attained by Teacher, Number of Years as a Teacher	.0934

4.4 Evaluation Measures of Alternative Imputation Methods

Tables 7, 8, 9, and 10 summarize evaluation measures produced for the multiple (five) imputations for items in Table 2 over 250 iterations. The best performing method is highlighted for each evaluation measure. Thus, as Table 7 shows, for Black Enrollment without administrative data (that is, Black Enrollment was imputed without the direct assignment of administrative data) predictive mean matching produced the best results: the mean relative deviation has the smallest value and the mean bias, standard deviation bias, quartile 1 bias, median bias, quartile 3 bias, and relative bias have the smallest absolute values. Based on the percentage of significant t-tests measure, the propensity score method outperformed the other three alternative methods. The majority of the evaluation measures identify predictive mean matching as the best performing, so this method is chosen as the method for imputing the actual missing values of the Black Enrollment item without direct assignment of administrative data.

Table 7: Evaluation Measures of Alternative Imputation Methods for Black Enrollment without Administrative Data

Method	MRD	Mean Bias	Std. Dev. Bias	% t-test sig.	Q1 Bias	Med. Bias	Q3 Bias	Rel. Bias
MCMC	5.67	19.08	-17.46	100.00	36.24	31.40	4.26	0.23
Propensity	8.67	-0.09	-87.27	0.84	33.30	41.02	11.50	0.00
PMM	0.33	-0.06	-2.35	6.30	0.63	0.23	0.88	0.00
Regression	5.67	19.09	-17.40	100.00	36.26	31.38	4.52	0.23

Table 8 displays the results for Black Enrollment with administrative data (that is, Black Enrollment was imputed with the direct assignment of CCD data and only the remaining missing records are imputed by alternative methods). Results for the best performing method are the same as the Black Enrollment without administrative data. As a result, predictive mean matching is chosen as the method for imputing the actual missing values of the Black Enrollment item with direct assignment of administrative data.

Table 8: Evaluation Measures of Alternative Imputation Methods for Black Enrollment with Administrative Data

Method	MRD	Mean Bias	Std. Dev. Bias	% t-test sig.	Q1 Bias	Med. Bias	Q3 Bias	Rel. Bias
MCMC	5.42	18.77	-17.01	75.50	34.49	29.11	4.59	0.26
Propensity	8.08	-3.73	-79.07	1.61	33.56	40.42	5.85	-0.01
PMM	0.29	0.02	-1.21	3.62	0.44	0.52	-0.75	0.00
Regression	5.38	18.53	-17.17	75.50	34.40	28.98	4.05	0.26

Table 9 displays the results for Newly Hired Teachers. The majority of the evaluation measures, with the exception of the percentage of significant t-tests measure and relative bias, identify predictive mean matching as the best performing method. Therefore, it is chosen as the imputation method for the real missing values.

Table 9: Evaluation Measures of Alternative Imputation Methods for Newly Hired Teachers

Method	MRD	Mean Bias	Std. Dev. Bias	% t-test sig.	Q1 Bias	Med. Bias	Q3 Bias	Rel. Bias
MCMC	1.06	1.20	-2.52	97.60	2.71	2.08	0.57	0.36
Propensity	0.82	-0.02	-2.93	3.20	1.62	0.98	-0.53	0.00
PMM	0.34	0.00	-0.77	4.80	0.35	0.45	-0.15	0.00
Regression	1.07	1.21	-2.51	97.60	2.71	2.08	0.57	0.36

Table 10 displays results for the Pension Check item. Some evaluation measures, such as the mean relative deviation, standard deviation bias, quartile 1 bias, median bias, and quartile 3 bias identify predictive mean matching as the best performing method, while other measures, such as the mean bias, percentage of significant t-tests, and relative bias identify the propensity score method as the best method. Based on the majority of the measures, the predictive mean matching is chosen as the imputation method for the real missing values.

Table 10: Evaluation Measures of Alternative Imputation Methods for Pension Check

Method	MRD	Mean Bias	Std. Dev. Bias	% t-test sig.
MCMC	10.46	4502.29	-13591.67	94.40
Propensity	7.94	-21.67	-11785.99	10.00
PMM	5.61	58.45	-214.95	20.80
Regression	10.49	4533.63	-13586.90	94.00

Table 10: (cont.)

Method	Q1 Bias	Med. Bias	Q3 Bias	Rel. Bias
MCMC	18478.03	6862.80	-4479.06	0.24
Propensity	12052.52	2236.33	-7400.39	0.00
PMM	559.16	-1835.51	1298.24	0.01
Regression	18502.78	6901.80	-4430.94	0.24

In addition to the evaluation measures listed above, the relative increase in variance is calculated to report the increase in variance due to multiple imputation. This measure quantifies the influence of the missing data on the sampling variance of a parameter estimate. The relative increase in variance is calculated by the following formula:

$$relative\ increase\ in\ variance = \frac{V_B + V_B/m}{V_W}$$

where V_B is the between-imputation variance, V_W is the within-imputation variance, and m is number of imputations. The denominator of this formula estimates the sampling variance that would have resulted had there been no missing data and the numerator quantifies the additional sampling variation that accrues from the missing data.

Table 11 shows the relative increase in variance for the alternative imputation methods. For Black Enrollment (without administrative data), Black Enrollment (with administrative data), and Newly Hired Teachers the relative increase in variance is negligible for every alternative method. For Pension Check imputed by the MCMC, propensity score, and regression methods the relative increase in variance is respectively

42.20%, 59.95%, and 41.33%. This means that the sampling fluctuation due to the missing data for each of the above mentioned methods is larger (by the corresponding percent) than the sampling variance of a complete-data analysis. Further research can address the reasons for small values of relative increase in variance.

Table 11: Relative increase in variance (%)

Method	Black Enrollment (without admin. data)	Black Enrollment (with admin. data)	Newly Hired Teachers	Pension Check
MCMC	0.44	0.03	2.02	42.20
Propensity	6.64	0.42	3.61	59.95
PMM	0.55	0.30	3.95	11.05
Regression	0.38	0.03	1.94	41.33

Tables 12, 13, 14, and 15 show the correlations between the response variables and their covariates. Generally, the correlations after the predictive mean matching method remained very similar to correlations of the observed data. In fact, some of the correlations between the covariates and the response variable improved after the predictive mean matching method compared to the observed data. The correlations after the hot deck method were generally lower than the correlations of the observed data.

Table 12: Correlation between Black Enrollment (without administrative data) and Covariates

	Covariates				
Group of Data	CCD Black Enrollment	Total Teachers	CCD Free / reduced lunch	Vice Principals	Number of Black teachers
Observed	0.94	0.40	0.49	0.46	0.73
PMM	0.93	0.46	0.41	0.43	0.74
Hot Deck	0.70	0.32	0.35	0.34	0.51

Table 13: Correlation between Black Enrollment (with administrative data) and Covariates

	Covariates				
Group of Data	CCD Black Enrollment	Total Teachers	CCD Free / reduced lunch	Vice Principals	Number of Black teachers
Observed	0.94	0.40	0.49	0.46	0.74
PMM	0.95	0.37	-0.20	0.54	0.61
Hot Deck	0.38	0.18	0.03	0.34	0.36

Table 14: Correlation between Newly Hired Teachers and Covariates

	Covariates					
Group of Data	Total Teachers	Vice Principals	Custodial/ Security	Black Enrollment	White Enrollment	Special needs
Observed	0.48	0.41	0.31	0.30	0.33	0.37
PMM	0.69	0.57	0.23	0.41	0.41	0.50
Hot Deck	0.56	0.45	0.19	0.16	0.46	0.37

Table 15: Correlation between Pension Check and Covariates

	Covariates	
Group of Data	**Highest degree**	**Number years as teacher**
Observed	0.13	0.29
PMM	0.31	0.36
Hot Deck	0.09	0.02

Table 16 below shows the results of the t-test for the means. The p-values indicate that the means of the alternatively imputed data and the observed data are significantly different. Therefore, the means are not preserved using the predictive mean matching method. The p-values also indicate that the means between the data imputed using hot deck and the observed data are not significantly different for Black Enrollment (with and without administrative data) and Pension Check. However, this is not the case for the Newly Hired Teachers item. Therefore, the means are preserved for Black Enrollment (with and without administrative data) and Pension Check, but not for the Newly Hired Teachers item. It is important to note that the hot deck values underwent consistency edits, while the predictive mean matching values did not.

Table 16: T-test for Means

Item	Reported		Imputed with PMM			Imputed with Hot Deck		
	Mean (Std. Dev.)	**N**	**Mean (Std. Dev.)**	**N**	**T-test for Means**	**Mean (Std. Dev.)**	**N**	**T-test for Means**
Black Enrollment (without admin. data)	85.10 (159.30)	7020	116.00 (196.00)	461	<.0001	99.09 (168.20)	461	.0687
Black Enrollment (with admin. data)	86.93 (162.20)	7448	149.90 (150.20)	33	.0261	89.30 (124.10)	33	.9332
Newly Hired Teachers	3.38 (4.37)	7245	4.56 (3.44)	236	<.0001	5.04 (6.07)	236	<.0001
Pension Check	19398.10 (18166.9)	670	14820.60 (16874.8)	235	.0005	20534.50 (22959.7)	233*	.4940

*Smaller than N for PMM by 2 records. The hot deck method did not impute these two values.

5. Conclusions and Further Research

At the record level, nearly all (96.5%) of the SASS school records were matched with a CCD record. However, at the survey item level, only nine items in Table 1 had acceptable rates of reported values on the CCD at 98%, with the remaining nineteen survey items having rates of reported values on the CCD of less than 42%. Among the nine items with high rates of reported values on the CCD, only Black Enrollment showed insignificant difference between SASS and CCD values. Therefore, Black Enrollment is the only potential candidate for replacement with CCD data.

The evaluation measures output from the truth deck simulation identified the predictive mean matching method as the best alternative imputation method for all four items in Table 2. However, the evaluation measures used to compare the predictive mean matching method to the hot deck method were not as clear. There is not enough evidence to conclude that the predictive mean matching method improved the quality of the imputation. The hot deck method preserved the means better, especially when using direct assignment of the CCD data first. Yet, the predictive mean matching method performed slightly better in preserving the correlations between the covariates and the response variables. Further research includes applying the consistency edits to the items in Table 2 after the predictive mean matching imputation has been performed. This would allow for a better comparison of the t-test for the means between the predictive mean matching method and the hot deck method. In the future, the alternative imputation methods will also be applied to more SASS items with different data types such as binary and categorical.

6. Limitations

The first limitation was the time lag between the CCD and SASS. SASS data was collected two years after the CCD data. This difference in time could have affected the results of the CCD quality measures. Another limitation was the fact that the current hot deck method could not be run on the simulated data. Hence, the distance to the true value between hot deck and PMM could not be measured at the record level. In addition, the hot deck data used in this study has undergone consistency edits whereas the data that was imputed using the predictive mean matching method has not. This may have allowed hot deck to better preserve the means. The final limitation was time and storage. The truth deck dataset process was limited to 250 iterations because the truth deck program took approximately six hours to run, and the output files were very large.

7. Software

The outputs/codes/data analysis for this paper were generated using SAS/STAT® software, Version 9.2. Copyright © 2002-2008 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

8. Acknowledgements

We would like to acknowledge and thank our reviewers Stephen Ash and T. Trang Nguyen with the U.S. Census Bureau. We would also like to acknowledge and thank Mei Li, Kelly Mathews, and Allison Zotti as members of the research team at the U.S. Census Bureau.

9. References

Durrant, G. 2005. Imputation Methods for Handling Item-Nonresponse in the Social Sciences: A Methodological Review. *ESRC National Centre for Research Methods and Southampton Statistical Sciences Research Institute (S3RI), University of Southampton*.

Grannell, A. and Murphy, H. 2011. Using Multiple Imputation to Adjust for Survey Non-Response. Retrieved from http://www.statistical-solutions-software.com/wp-content/uploads/2012/07/Multiple-Imputation_ASC2011.pdf

Hippel, P. 2012. Imputing Skewed Continuous Variables Under a Modified Normal Model: A Comparison of Methods. *Sociological Methods and Research*.

Little, R.J.A. and Rubin, D.B. 1987. *Statistical Analysis with Missing Data*, John Wiley & Sons.

Little, R. 1988. Missing-Data Adjustments in Large Surveys. *Journal of Business and Economic Statistics* (6). 287-296.

National Center for Education Statistics. *Schools and Staffing Survey (SASS) Overview*. Retrieved December, 2013 from http://nces.ed.gov/surveys/sass/overview.asp.

Rosenbaum, P.R. and Rubin, D.B. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* (701). 41-55.

Rubin, D.B. 1976, Inference and Missing Data. *Biometrika* (63). 581-592.

Rubin, D.B. 1987, *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley & Sons, Inc.

Rubin, D.B. 1988. An Overview of Multiple Imputation. In *JSM Proceedings*, Survey Research Section. Alexandria, VA: American Statistical Association. 79-84.

Schafer, J.L. 1997. *Analysis of Incomplete Multivariate Data*, New York: Chapman and Hall.

Schenker, N., Raghunathan, T.E., Chiu, P., Makuc, D.M., Zhang, G., and Cohen, A.J. 2006. Multiple Imputation of Missing Income Data in the National Health Interview Survey. *Journal of the American Statistical Association* (475). 924-933.

Ziegelmeyer, M. 2011. Illuminate the Unknown: Evaluation of Imputation Procedures Based on the SAVE survey. *Advances in Statistical Analysis* (97). 16-18.