Analysis of Multiple Imputation Techniques for the Survey of Local Election Officials

Timothy Markham[1], Eric Falk[1], David McGrath[1]
Jeffrey Schneider[1]
[1]Defense Manpower Data Center, 1600 Wilson Boulevard Suite 400, Arlington, VA 22209

Abstract
Survey research commonly encounters the problem of analyzing data that contains incomplete or missing information. Acock (2005) notes that this missing information can produce biased estimates, distort statistical power, and allow researchers to draw invalid conclusions. Fortunately, these risks can be effectively curbed in many situations with imputation. The imputation theory, developed by Rubin in 1987, has been proven to successfully estimate a parameter of interest, as well as accurately assess the variability of the estimate. The purpose of this paper is to test the use and applicability of the multiple imputation method PROC MI in SAS as a valid and robust missing data technique as compared to a multiple hot deck process. The Department of Defense Manpower Data Center (DMDC) applied these two imputation methods to the Federal Voting Assistance Program (FVAP) survey of Local Election Officials (LEOs) and analyzed their efficacy. The failure of the LEO survey data to meet the assumptions of PROC MI makes that method intractable. While the multiple hot deck method provides more plausible results, it struggles to incorporate complex logical relationships between questions.

Key Words: PROC MI, hot deck, multiple imputation, voting

1. Background

FVAP is the agency within the Department of Defense responsible for assisting uniformed service members and other overseas citizens in voting. To monitor the number of voters and the efficiency of FVAP's programs, DMDC administers six surveys after each national election. The most recent round of surveys was fielded following the November 2010 federal elections, focusing on uniformed service members, military spouses, Unit Voting Assistance Officers, Department of State Voting Assistance Officers, overseas civilians, and LEOs, respectively.

A LEO is a person who, either as an individual or as a part of a group, oversees the election process at the jurisdiction level. Currently, there are 7,296 jurisdictions in the United States and U.S. territories. While many of these are counties, some states, such as Wisconsin, track ballots at the sub-county level. Other states, such as Maine and Alaska, monitor absentee ballots at the state level.

One of the functions of a LEO is to monitor the absentee ballot process. As a result, DMDC's survey of LEOs attempts to quantify the number of ballots transmitted, received, and counted in each jurisdiction so that national totals can be estimated. The majority of the survey questions focus on voters who are subject to the Uniformed and Overseas Citizens Absentee Voting Act (UOCAVA) of 1986, which ensures that qualifying U.S. citizens can register and vote absentee for federal offices. Questions on the survey range from the number of UOCAVA-covered registered voters to the number of ballots from UOCAVA-covered voters rejected due to problems such as a lack of signature. Questions with numeric responses have sub-items for uniformed service members, overseas citizens, and totals. The survey instrument is available on FVAP's website, http://www.fvap.gov.

DMDC's survey of LEOs after the November 2010 elections was a census of all 7,296 jurisdictions that was administered with paper and web questionnaires. For a jurisdiction to be considered a complete and eligible respondent, it had to respond to at least one question. 3,894 jurisdictions returned complete and eligible surveys, making the response rate 53%. The rate was calculated according to the RR3 recommendation of the American Association for Public Opinion Research (2008). For weighting, the jurisdictions were post stratified by the number of registered voters, an administrative variable, as shown in Table 1.

Table 1: Counts and Final Weight by Poststratum

Poststratum	*Number of Registered Voters*	*Jurisdiction Count (A)*	*Complete Eligible Cases (B)*	*Final Weight (A/B)*
1	> 5,001	4,200	2,167	1.94
2	5,001–10,000	829	436	1.90
3	10,001–29,202[a]	1,267	663	1.91
4	29,203–40,000	237	136	1.74
5	40,001–75,000	319	197	1.62
6	75,001–100,000	102	54	1.89
7	100,001–20,000	162	112	1.45
8	200,001–360,000	84	66	1.27
9	< 360,000	96	63	1.52

[a] To encourage response from large jurisdictions, FVAP called the largest 1,000 jurisdictions, which included all jurisdictions with more than 29,202 registered voters. To capture the effect of these calls on response propensity, the poststrata were created so that none of the largest 1,000 jurisdictions is in a poststratum with a jurisdiction that did not receive a call.

Missing data poses an important problem in DMDC's LEO survey because the goal was to estimate national totals of UOCAVA-covered voters. If values are not imputed for missing data, totals would be underestimated. In essence, a missing value is not taken into account when weighted totals are created because a value that does not exist cannot be weighted. Therefore, any jurisdiction who left a question blank when the actual value was greater than zero, for example because the numbers were not readily available when the LEO completed the survey, would be contributing zero to the national total when it should be contributing a positive number.

Missing data occurred widely in the LEO survey, ranging from 3 to 48 percent per numeric item. This means that of the responding jurisdictions, nearly half did not respond to some of the questions. Large jurisdictions responded at significantly lower rates than small jurisdictions, as seen in Table 2. Those LEOs that responded by paper questionnaire were more likely to respond to each item than those that filled out the survey on the web, 15 to 20 percent more likely in almost all of the 275 numeric items. Due to the large amount of missing data, imputation was necessary in order to obtain reasonable national estimates for many of the numeric items.

Table 2: Average Response Rate per Numeric Item by Poststratum

Poststratum	*Number of Registered Voters*	*Percent Responding*
1	> 5,001	83%
2	5,001–10,000	66%
3	10,001–29,202[a]	53%
4	29,203–40,000	50%
5	40,001–75,000	40%
6	75,001–100,000	41%
7	100,001–20,000	38%
8	200,001–360,000	33%
9	< 360,000	34%

[a] To encourage response from large jurisdictions, FVAP called the largest 1,000 jurisdictions, which included all jurisdictions with more than 29,202 registered voters. To capture the effect of these calls on response propensity, the poststrata were created so that none of the largest 1,000 jurisdictions is in a poststratum with a jurisdiction that did not receive a call.

2. Imputation Considerations

Techniques for handling item nonresponse have been studied extensively over the past thirty years. Much debate has been sparked determining which methods to use are optimal. According to Acock (2005), methods that have deficiencies or are used incorrectly can lead to biased estimates, distorted statistical power and invalid conclusions. Among the most researched methods are the stochastic imputation works by Rubin as well as the hot deck approach by Ford.

Dempster, Laird, and Rubin (1977) and Rubin (1987) have proposed that multiple imputation can provide valid inference for item nonresponse. Multiple imputation is a technique where missing values are “filled in” or replaced by sets of simulated values to create multiple completed datasets. The multiple completed datasets could now be analyzed by themselves using standard statistical methods. The datasets could then be combined, which would take the stochastic processes into account. Rubin has proposed that multiple imputation requires somewhere between three to ten simulations to provide sufficient results.

More specifically, Rubin describes multiple imputation through Bayesian methods. As such, a parametric model should be specified for complete data under missing at random, and a prior distribution for the unknown model should be assumed. Then multiple independent draws should be simulated from the conditional distribution of missing values. Following his initial paper, Rubin (1987, 1996) furthered his belief that multiple imputation should be the primary method when dealing with missing data. Rubin noted that other missing data techniques often require subject matter expertise and often do not provide scientifically robust estimates.

Prior to Rubin's work, Ford (1983) proposed the hot deck imputation method, which uses data from the data set to impute the missing values. Hot deck imputation takes data from usable cases, called donors, and matches them with suitable missing data. The exact matching technique can vary based on situation. An application of multiple imputation and hot deck methods culminated with Fay (1992).

Multiple imputation software is discussed in Horton and Lipsitz (2001). The SAS Institute released their multiple imputation procedures called PROC MI and PROC MIANALYZE. PROC MI's default method for imputation of missing data is the Markov Chain Monte Carlo (MCMC) algorithm. As Allison (2005) notes, this method is based on the assumption of multivariate normality which implies valid imputations may be generated by linear regressions. PROC MI is considered to be computationally efficient and has the capacity to handle arbitrary patterns of missing data.

3. Methodology

3.1 Data Editing

Once the survey field period was closed, a number of steps had to be taken to prepare the data before either PROC MI or the hot deck imputations could commence. First, a series of data edits was performed. For all numeric data points, if a jurisdiction provided values for both the uniformed service member and overseas citizen sub-items but not the total, or if the total was not equal to the sum of the other sub-items, the total was set equal to the sum of those sub-items. Also, data edits were performed to ensure that the data conformed to the logical relationships between questions. For example, the number of regular UOCAVA absentee ballots received by a jurisdiction must be lower than the number of ballots transmitted by that jurisdiction. If the reported value of ballots received for any sub-item was higher than the reported value of ballots transmitted for the corresponding sub-item, the former was set equal to the latter.

For jurisdictions that responded to both the total and one other sub-item of a question, the other sub-item was set equal to the difference between the total and the reported sub-item. If, for instance, a jurisdiction reported that it received 10 regular UOCAVA absentee ballots from uniformed service members and 15 total regular UOCAVA absentee ballots, the number of ballots received from overseas citizens was changed from missing to five. This imputation was deterministic, as five is the only number that can preserve the logical relationship between sub-items in this example.

In some cases, jurisdictions provided totals for some questions but information for neither uniformed service members nor overseas citizens. Instead of using PROC MI or hot deck imputation for these cases, a weighted sum of the responses of jurisdictions that provided complete data for all sub-items was created and proportions of the total were assigned to uniformed service members and overseas citizens. For the jurisdictions that provided only a total, these proportions were then multiplied by that total value and the results replaced the missing data. For nearly all questions, uniformed service members contributed 60 to 80 percent of the total based on jurisdictions responding to all sub-items. If 75 percent of ballots received from jurisdictions that completely responded to that question came from uniformed service members, for example, 75 percent of the total was imputed into that sub-item for jurisdictions that provided only a total.

To this point, none of the imputations were truly stochastic. In other words, they depended on other observed values from the same jurisdiction. For questions in which a jurisdiction answered no part, however, the previous methods could not be applied. Instead, two approaches were taken and compared using SAS software: PROC MI and multiple hot deck imputation.

3.2 PROC MI

Schafer (1997) notes that the MCMC method, the default of PROC MI, is used to fill in missing values for arbitrary missing patterns. The non-monotone missingness within the LEO survey data made the MCMC method the most viable. Five sets of imputations were created to account for the uncertainty of imputed values. When no other parameters were used with PROC MI, many of the imputed values were negative. However, because all numeric items deal with the absentee ballot process, negative numbers are not logical. As a result, both five unbounded imputations and five imputations with the PROC MI option MIN=0 were developed. MIN=0 ensures that no imputed value is negative. In both cases, values were only imputed for the uniformed service member and overseas citizens sub-items. The sum of these values was then imputed for the total to preserve the logical relationship. Once the totals were created, each group of five imputations was then analyzed using PROC MIANALYZE to get total and variance estimates. For totals, averages of the five imputations were taken to be the point estimate for each jurisdiction. The total variance is defined as follows, where I_i is the i^{th} imputation and *Var(Var(I))* is the variance between the individual imputation variances:

$$Total\ Variance = \sum_{i=1}^{5} Var(I_i) + \left(1 + \frac{1}{5}\right) * Var\big(Var(I)\big)$$

3.3 Multiple Hot Deck

As an alternative to PROC MI, a multiple imputation system was employed using a SAS macro for hot deck imputation created by Iannacchione (1982) and adapted for iterative use by Ellis (2007). The macro itself produces one complete data set based on a weighted hot deck technique, which completes imputations so that the complete data set maintains the weighted distribution of actual responses. Ellis states that weighted hot deck is designed for data sets with at least 10 percent missingness, whereas unweighted, which does not account for the distribution of data, is for less than five percent missingness. Either version can be used if the level of missing data is between five and 10 percent.

For this hot deck process, numeric items were divided into groups of related questions with the aim of preserving the logical relationships between these items. For example, the numbers of regular UOCAVA ballots transmitted, received, and counted were imputed into a jurisdiction that had not responded to those items from the same complete donor case.

To attempt to keep donor cases and recipient jurisdictions similar, jurisdictions were divided into cells. The variables that determined the cells changed based on the number of donor cases available for a particular group of questions. For all groups of questions, jurisdiction size as delineated in the poststrata was used in the definition of cells. In groups of questions with enough complete cases to become donors, jurisdiction type was also used. For this purpose, type was collapsed into county and state and sub-county, which consists of villages, towns, and townships. Region was used if jurisdiction size and type created cells with too few donor jurisdictions, as each jurisdiction could only be used three times in an imputation by the design of the macro. The regions include Northeast, Midwest, South, and West. In some groups of questions, too few complete cases existed to include either jurisdiction type or region in the cell definition. Jurisdiction size was the only variable used in these instances.

As in the case with PROC MI, single imputations do not properly account for the variance of imputed values. To account for the level of uncertainty associated with imputation, five data sets were created, as in the case of PROC MI. Ellis' macro was embedded in another macro, which ran the hot deck five times and compiled the resulting imputations into one data set of the same format as the outputted PROC MI data set. The complete multiple imputation data set was then analyzed using PROC MIANALYZE so that the results could be compared to the PROC MI data set. Totals and variance were found using the same definitions as in the PROC MI case.

4. Results

4.1 Weighted Totals

The weighted point estimates, minimum observation, and maximum observation for the total number of regular UOCAVA absentee ballots received using each multiple imputation method can be found in Table 3. For the unbounded PROC MI data set, the minimum value is negative. In fact, 105 observations, or 2.6 percent, have negative values after the five imputations have been averaged. These numbers are illogical in the context of the survey. Also, PROC MI does not account for logical relationships between questions on either an observation or a total level. For example, the weighted total for the number of regular UOCAVA absentee ballots submitted for counting is 135,791 and the weighted total for the number counted is 133,464. This means that, according to this data, 102 percent of ballots submitted for counting were actually counted, which cannot be true.

While using the MIN option on PROC MI prevents negative imputed values, this process has problems of its own in the context of the LEO survey. First, the logical relationships between questions still cannot be controlled, resulting in similar analytic problems as the unbounded PROC MI data set. Also, the assumption of PROC MI is that the data follows a multivariate normal distribution. However, the survey responses show that this is not the case. For some questions, as much as 95 percent of survey responses were zero. PROC MI attempts to re-impose a normal distribution by imputing values that are skewed to the right. As a result, totals and means are significantly higher than either the unimputed data or the unbounded PROC MI, as much as 30 percent higher.

The multiple hot deck imputation, by only imputing values reported by other jurisdictions, prevents imputed values from being negative. In cases where a jurisdiction did not respond to an entire group of questions that was imputed together, logical relationships were maintained. However, in many instances jurisdictions responded to some but not all of the questions that were imputed together. In these cases, logical relationships could not be taken into account, as only some of the values were taken from the donor jurisdictions. This problem was compounded by illogical survey responses that could not be fixed because all questions had not been answered. For instance, if a jurisdiction reported transmitting 100 regular absentee ballots and counting 200 but did not report a number of ballots received, these values were not corrected and any imputed value necessarily cannot conform to all logical relationships. As a result, totals to individual questions appear more statistically valid than either of the PROC MI cases, but ratios of the questions do not make analytic sense. The weighted estimate of regular absentee ballots submitted for counting is 183,850 and the estimate of ballots counted is 184,242, meaning that over 100 percent of ballots were counted.

Table 3: "Of the total number of regular absentee ballots that your jurisdiction transmitted to UOCAVA voters for the November 2010 general election, how many were returned [total]?"

Imputation Technique	*Weighted Total*	*Minimum*	*Maximum*
PROC MI, no min	142,394	-208	5,144
PROC MI, MIN=0	228,153	0	5,144
Hot Deck	193,661	0	5,144

4.2 Variance Estimation

The purpose of multiple imputation is to inflate variance to account for the uncertainty of imputed values. The variances of the three procedures associated with the estimates in Table 3 can be found in Table 4. While the hot deck data set has the highest variance, it also has the smallest relative increase in variance, which comes from the variance between the individual imputations, at under 0.5 percent.

Table 4: Variances Associated with Table 3

Imputation Technique	*Variance Within*	*Variance Between*	*Total Variance*	*Relative Increase in Variance*
PROC MI, no min	12.79	0.29	13.14	2.73%
PROC MI, MIN=0	13.15	0.79	14.10	7.24%
Hot Deck	18.31	0.07	18.39	0.43%

5. Summary and Conclusions

Using PROC MI for the LEO survey is intractable, whether the procedure is unbounded or uses a MIN=0 option. The complex logical relationships between questions in the LEO survey make an imputation scheme like PROC MI undesirable because it cannot account for them, but the biggest problem is the assumption of multivariate normality. With data that is highly skewed and so many reported values of zero, the normality assumption breaks down and the weighted estimates are highly inaccurate.

Because hot deck only imputes values that have been reported by other jurisdictions and cells can be created to limit possible donor jurisdictions based on characteristics such as jurisdiction size and type, this method creates totals that are more plausibly analytically. However, two problems exist within hot deck for this survey. First, the process cannot handle the numerical logic of the questions for jurisdictions that responded to some but not all the question in an imputation group. Second, the administrative variables available to create the cells are not ideal. Nearly all numeric items in the LEO survey deal with UOCAVA voters. However, administrative data concerning the number of UOCAVA voters in a jurisdiction does not exist, and the number of UOCAVA voters is not accurately predicted by overall jurisdiction size.

5.1 Future Research

The LEO survey is conducted biannually, so the issue of missing data within the context of this survey will be revisited again after the national elections in 2012. In order to create the most accurate estimates possible, DMDC would like to do further research on multiple imputation for the 2010 survey. An empirical comparison between weighted and unweighted hot deck procedures could show the value of the weighted version or could indicate that groups of questions should be imputed using different methods based on the percent of missingness for each group. Also, DMDC would like to find a more efficient means of dealing with the complex logical relationships between questions, perhaps with changes to the survey instrument. Developing questions that are independent of other survey responses, do not result in as many zero responses, and encourage response from large jurisdictions or jurisdictions with large UOCAVA populations could improve the efficacy of estimates in 2012.

Other procedures for handling multiple imputation that could not be used for analysis of DMDC's 2010 survey due to time constraints may also be tested prior to the implementation of the 2012 iteration of the LEO survey. For example, the Bayesian bootstrap method, which simulates the posterior distribution of model parameters, may be an appropriate alternative. Also, IVEware is software that can be launched from SAS to perform multiple imputation using a sequential regression imputation method and could prove to be a useful tool in both creating and analyzing multiply imputed data sets.

References

Acock, A.C. (2005). Working with missing values. *Journal of Marriage and Family*, 67, 1012-1028.

Allison, P.D. (2005). Imputing categorical data with PROC MI. *Proceedings of the Thirtieth Annual SAS Users Group International Conference* (Paper 113-30). Cary, NC: SAS Institute, Inc.

American Association for Public Opinion Research. (2011). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 7th Edition. Deerfield, IL: AAPOR.

Dempster, A.P., N.M. Laird, and D.B. Rubin. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1-38.

Ellis, B. (2007). A consolidated macro for iterative hot deck imputation. *Northeast SAS Users Group Inc. 20th Annual Conference Proceedings* (PO03). Baltimore, MD: NESUG.

Fay, R.E. (1992). When are inferences from multiple imputation valid? *Proceedings of the Section on Survey Research Methods* (pp. 227-232). Alexandria, VA: American Statistical Association.

Ford, B. L. (1983). An overview of hot-deck procedures. *Incomplete data in sample surveys*, 2, 185-207.

Iannacchione, V.G. (1982). Weighted sequential hot deck imputation macros. *Proceedings of the SAS Users Group International Conference* (pp. 759-763). San Francisco, CA: SUGI.

Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Rubin, D.B. (1996) Multiple imputation after 18+ years (with discussion). *Journal of the American Statistical Association*, 91, 473-489.

Horton, N.J. and S.R. Lipsitz. (1996). Multiple imputation in practice: comparison of software packages for regression models with missing variables. *The American Statistician*, 55, 244–254.

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

“Uniformed and Overseas Citizens Absentee Voting Act.” Pub. L. no. 99-410. 100 Stat. 924 (1986).