Missing Data Methods for the CCM Component Error Estimation

Vincent Thomas Mule, Donald Malec, Lynn Imel, Nganha Nguyen and Michael Moldoff
U.S. Census Bureau

Keywords: Census, Coverage Measurement

Abstract: In the 2010, The Census Coverage Measurement (CCM) will estimate the components of census coverage in addition to net error. This includes estimates of correct and erroneous enumerations in the Census. This paper focuses on missing data methods examined for handling census records in the estimation with unresolved enumeration status who also did not report name and two characteristics during their enumeration.

I. Introduction

One of the new goals for the Census Coverage Measurement (CCM) program is to estimate the component of census coverage errors. Whitford (2008) provides a high-level background on the proposed coverage measurement estimates for component error that includes erroneous enumerations and omissions. The CCM will only be tallying the number of whole-person census imputations and will not evaluate their correctness. For the remaining person in housing unit records in the 2010 Census, the CCM will estimate the number that were either correct or erroneous.

In order to estimate the number of correct or erroneous enumerations for components, the CCM program had to expand the matching operations beyond what was done in the past for dual system estimation (DSE). To reduce matching error when implementing the DSE, one of the requirements for a case to be a correct enumeration in the Enumeration (E) sample[1] is completeness. This requires that the E-sample case have a reported name and at least two other reported characteristics. Any cases that did not meet this requirement were determined to have Insufficient Information for Matching and Followup[2] and treated as erroneous enumerations for DSE. For the 2010 CCM, these cases are being called "Insufficient Information for DSE processing" since cases can and will be handled differently for net error and component error estimation. The focus of this paper is on the handling of these cases for component estimation. Mule (2008) provides some details on how the census records are classified for net and component estimation based on reported information.

For component estimation, the CCM is doing a couple of things differently than for net error. First, we are relaxing the requirement of completeness. The CCM will estimate

---

[1] The Enumeration sample is a representative sample of data-defined census enumerations. A data-defined enumeration has two characteristics reported in the Census.
[2] These cases are also known by their net error match code of "KE".

the number that were correct or erroneous for the universe of data-defined cases in Census housing units. Second, we are expanding the geographic area for being classified as correct to include the entire nation. Mule (2008) provides details on the estimation methodology used to generate these estimates.

To support component error, we will attempt to clerically match the Insufficient Information cases and use the results in component estimation. Livermore Auer (2005) documented a study designed to clerically match cases deemed insufficient information for matching and followup to the Population (P) sample using the 2000 Accuracy and Coverage Evaluation (A.C.E.) data. The study examined how person interview and followup information for the P-sample cases could be used to try to determine the enumeration status of the Insufficient Information for DSE processing cases. His results showed that approximately one-half of these cases were able to be assigned an enumeration status for component error estimation. Moldoff (2008) documents how the CCM clerical matching operation has been expanded to attempt to clerically match these cases.

Table 1 summarizes the possible results of the E sample for component estimation after clerical matching is completed. The matching will be able to resolve the status for some of the Sufficient and Insufficient cases. There will be some Sufficient Information cases that will be unresolved. The CCM program has experience imputing these unresolved cases for dual system estimation that can be drawn upon. These cases are eligible for matching and followup so we can utilize similar covariates like Before Followup and After Followup information in making the imputation cells. This leaves the Insufficient Information cases that are unresolved.

Table 1: Enumeration Status of Cases for Component Error Estimation

	Information Status for DSE Processing	
	Sufficient	Insufficient
Resolved		
Unresolved		

In documenting the study, Livermore Auer stated this conclusion for the unresolved rate of the Insufficient Information cases in his analysis:

> There is a high unresolved rate among KE records as many of these cases did not match or matched with low confidence and were not sent to followup. It may be assumed that a similar rate will be observed in the future and since most of these records do not have a discernable name they will not be followed up. Appropriate missing data procedures will have to be applied to these cases.

Based on this matching research of 2000 data, missing data methods were examined to see how they accounted for unresolved enumeration status cases. We examined missing data methods using insufficient information cases from the 2006 CCM Test. Since the Insufficient Information for DSE processing cases is a new source of missing data, we researched several missing data procedures to see what was appropriate.

Section 2 provides background on the 2006 CCM Test in Travis County, Texas and Cheyenne River Sioux Reservation in South Dakota. Section 3 provides details on the missing data assumptions in this analysis. Section 4 presents results based on these assumptions using the 2006 CCM test data. Section 5 presents some preliminary conclusions and areas where we will be conducting additional research.

II. Background on the 2006 CCM Test

This research was done on CCM data from the 2006 Census Test in Texas and Cheyenne Sioux Reservation. This was the first test of the CCM interviewing and matching being conducted to support estimating the component of census coverage error.

A. Limitations of the 2006 CCM Test

Since this was a census test, this analysis is subject to multiple limitations including, but not limited to:

- it was not possible to obtain the level of cooperation usually obtained in a census environment.

- the unresolved rates were considerably higher than would be expected for 2010.

- the test covered only a limited area so it was not possible to search for duplicate records outside the site.

- many operations were being implemented for the first time and will need refinement for Census 2010; there may be additional non-sampling errors.

- some cases were sent to followup for evaluation purposes that would have been considered to have been resolved. This action led to some cases being classified as unresolved in this analysis because of an unsuccessful followup attempt.

- CCM listing or housing unit operations were not a part of the 2006 Census Test. This may have impacted the rates of resolved and unresolved cases. The resolution of the Insufficient Information cases for component estimation is improved when a personal interview is conducted at the housing unit of the person. The 2010 sample design tries to achieve this by using the results of the initial housing unit matching to identify housing units on the census list that were not listed by the CCM. These will not be part of the P sample for dual system estimation but the interview results can be used in the matching to resolve E-sample cases.

- E-sample cases with unresolved duplicate links outside of the block cluster search area were treated as not having duplicates. An examination led to the conclusion

that a vast majority of these people were linked on common names only and were not the same person. The 2010 component missing data model will be developed to account for unresolved duplicate links since the search will include the entire nation.

B. 2006 CCM Test Data

Table 2 summarizes the enumeration status of the E sample. Based on the results of the clerical matching operation, a case was determined to be correct for component error estimation if it was enumerated a) only once in the test site or b) if the person was enumerated more than once and this is the correct area[3] where the person should have been enumerated on Census Day. Cases determined to be fictitious, born after Census Day, died before Census Day or a duplicate not in the correct area are resolved as erroneous enumerations. Since we have a binary outcome, we will focus on the erroneous enumerations.

Table 2 shows that 63.4 percent of the Insufficient Information for DSE processing cases had an unresolved enumeration status. As forecast by Livermore Auer, a high unresolved rate for these cases has been observed. This is higher than the 15.0 percent of the Sufficient Information for DSE processing cases that had an unresolved enumeration status.

Examining the resolved cases for the two groups shows a difference. The erroneous enumeration rate for Sufficient Information cases is 1.9 percent (1.6 percent out of 85.0 percent) as compared to an erroneous enumeration rate of 7.3 percent (2.7 percent out of 36.6 percent) for the insufficient resolved cases.

Table 2: E-sample Enumeration Status for 2006 CCM Test Component Estimation

	Sufficient Information for DSE Processing			Insufficient Information for DSE Processing		
	Count	Weighted Total	Percentage	Count	Weighted Total	Percentage
Correct Enumeration	8,225	314,690	83.4%	256	11,760	33.9%
Erroneous Enumeration	174	6,171	1.6%	25	927	2.7%
Unresolved	1,486	56,675	15.0%	540	21,962	63.4%
Total	9,885	377,536	100%	821	34,648	100%

---

[3] Search area is the sample block cluster, one or more contiguous collection blocks, and the one ring of surrounding census collection blocks.

III. Missing Data Assumptions of Enumeration Status for 2006 Component Research

In this preliminary analysis, we made two assumptions about the missing mechanism for the unresolved cases. First, we assumed that the data were Missing At Random (MAR). Second, we assumed that the data were Missing Not At Random (MNAR).

A. Missing At Random Assumption

For MAR, this implies that given the observed data, the missingness mechanism does not depend on the unobserved data. The missing value mechanism can be expressed solely in terms of data that were observed. There is information available to use as a covariate so that by conditioning on that information makes the data MAR and leads to valid estimates. The CCM makes this type of assumption for missing enumeration status for E-sample data in the DSE. This is also referred to as an ignorable missing data situation.

For the Sufficient Information cases, we treat the unresolved enumeration status as item nonresponse. When imputing for unresolved enumeration status as item nonresponse with the MAR assumption, we used Imputation Cell Estimation. First, all people are placed in cells formed using relevant operational characteristics. The weighted proportion determined to be erroneous in each cell was calculated from the people in the cell with resolved status, and this proportion was assigned as a probability to each person in the cell with unresolved status. This was the same methodology used in the 2000 A.C.E. (Cantwell et al. 2001, Beaghen and Sands 2003). Since we want to account for the variability of this imputation in the overall variance estimation, this mean cell approach can be utilized by methods laid out in Yung and Rao (2000) for jackknife replication with post-stratification[4]. While there are other methods like Multiple Imputation that could have been considered, the CCM staff has experience with implementing replication methods in previous coverage surveys.

For the Insufficient Information cases, we examined treating the unresolved enumeration status cases as either *item* or *total* nonresponse. When treated as item nonresponse then we used covariates that were available to all cases. The available covariates include Census processing information and any information obtained during the matching operation before the person followup operation. The 1990 PES and 2000 A.C.E. both used Before Followup information about the case as an important covariate in their missing data models.

Beaghen and Sands (2003) determined for A.C.E. Revision II that utilizing followup information was the single most important improvement in the missing data methodology. Since most of the Insufficient Information for DSE processing cases[5] could not go to followup, we explored treating the unresolved Insufficient Information cases as total nonresponse. Since minimal information was collected about the person,

---

[4] Post-stratification referenced here is the typical ratio-adjustment to known control totals used in survey weighting. It is not the groupings of cases used in Dual System Estimation to reduce heterogeneity.

[5] Cases with a complete valid name but less than two reported characteristics were eligible for followup.

the assumption of total nonresponse is reasonable. With this total nonresponse assumption, we investigated a two-step process.

First, the unresolved insufficient cases were accounted for by a weighting cell adjustment. The weight adjustment method puts the cases into groups based on auxiliary information about the survey respondents. The grouping covariate is strongly related to the correct or erroneous outcome and also impacts the ability to resolve the insufficient cases. For the groups formed, the weights of the unresolved Insufficient Information cases will be spread to the other sample cases in the group. The weights will be allocated to the three other cells shown in Table 1 (this includes the Sufficient Unresolved cell).

After the weight adjustment has been performed, the remaining unresolved cases are those with Sufficient Information. For the second step, imputation cell estimation was used. This accounts for the item nonresponse of not being able to determine the status of these cases. The unresolved Sufficient Information cases were eligible for followup so after followup information could be used to form imputation cells.

B. Missing Not At Random Assumption

The second assumption was the data were Missing Not at Random. This is also referred to as a non-ignorable missing data situation. Even after conditioning on available information, the reason for observations being missing still depends on the unseen observations. To obtain valid inference, a joint model of the data and the missing mechanism is needed. This requires a determination of what is the appropriate model for the missing data mechanism. Since unresolved enumeration status of Insufficient Information for DSE processing cases is a new source of missing data, we did some preliminary analysis using this assumption.

Our Missing Not At Random models all fall into the framework of Little and Rubin (2002), Section 15.7, *Nonignorable Models for Categorical Data*. As suggested in that section, the EM algorithm is used to obtain maximum likelihood estimates of all parameters, which are then used to make mean imputations for the unresolved cells. We weighted the likelihood using the sampling weights.

C. Covariates

In this section, we list the covariates that were utilized in this preliminary analysis of missing data methods for component estimation.

Before Followup Groups

Previous coverage measurement surveys have used groupings of cases based on their status during the clerical matching operation before the person followup operation occurs. This status can be based on results for the individual person and also if other people in the housing unit have been matched to a person collected in the person interview. This was used in the 1990 Post-Enumeration Survey (Belin et al. 1993), 2000 A.C.E.

(Cantwell et al. 2001) and in net error pseudo-estimation[6] using 2006 CCM data (Seiss and Kilmer 2008). One possible solution for our component missing data problem is to apply the same covariates like this one used in the net error missing data.

The six Before Followup groups for the 2006 CCM were:

1. Match, No Followup
2. Match, Followup
3. Possible Match
4. Nonmatch, Other Persons in Housing Unit Match
5. Nonmatch, Whole Household Nonmatch
6. Duplicate/Potential Duplicate or Fictitious/Potential Fictitious

We examined some descriptive statistics of the sample cases by the six Before Followup groups. The Insufficient Information for DSE processing cases fell mostly into two groups. First, approximately 40 percent (13,892 out of 34,648) of the cases were determined to be in the *Match, No Followup* group. We also saw that 18.1 percent of the insufficient cases in this group were unresolved. Second, approximately, 44 percent (15,087 out of 34,648) of the cases were determined to be in the *Nonmatch, Whole Household Nonmatch* group. For this group, we saw that 98.1 percent of the insufficient cases were unresolved.

Sufficiency Information for DSE Processing Indicator

Since the insufficient information cases have a) a higher unresolved rate and b) a higher erroneous enumeration rate for the resolved cases, the missing data adjustment mechanisms may want to take advantage of this covariate whenever possible. Since the ratio of resolved cases-to-unresolved cases is approximately 1-to-2, it does have some drawbacks especially for Missing At Random methods. Table 2, shown earlier, provides some descriptive statistics for this covariate.

Type of Return for the Census Enumeration

There are several different ways that a person could have been enumerated during the 2006 Census Test. For this analysis, they have been classified into three groupings based on whether it was a) self or proxy reporting and b) whether it was a mail return or enumerator return. All mail returns are classified as self-reporting.

The three groupings of type of response in our analysis are:

1. Self-reported mail return
2. Self-reported enumerator return
3. Proxy-reported enumerator return

---

[6] Pseudo-estimation was the test implementation of estimation approaches using the CCM data in the 2006 Census test. Since this was not an official evaluation and there are several limitations in the test, these were called "pseudo-estimates."

Person Followup Question: Did he/she stay here all the time, move or go back and forth between two or more places?

The A.C.E. Revision II concluded that using information collected during person followup operations was able to produce more discriminating groups of whether people were correct or erroneous for the dual system estimates. In our analysis, we used one of the questions that provides a good indication of whether the cases in the group may be correct or erroneous. People who respond that they live here all the time are more likely to be correct as compared to the sample cases in the other two groups who are more mobile.

Since this question could only be asked of the sufficient cases that went out to followup, this variable will be used in the Missing At Random situations after a weighting adjustment applied to account for the insufficient unresolved cases. The insufficient cases are treated as total nonresponse in this adjustment. After the weighting adjustment, this will leave as unresolved only the sufficient cases that could not be determined to be correct or erroneous. All sufficient cases did not have to go to followup or did not provide an answer. So if the person did not answer this question, then other covariates could be used for that person.

IV. Different Missing Data Assumptions

Based on the missing data mechanism and the covariates selected in this preliminary research, we examined the performance of six missing data models.

1. *Missing At Random conditional on Before Followup Group only*

This missing data model examines the results if we apply an ignorable assumption using Before Followup group as the covariate. A further assumption is that enumeration status is independent of sufficiency status, given the Before Followup group. This covariate has been identified as beneficial when accounting for the missing data of Sufficient Information for DSE processing cases in net error estimation. The first approach shows the results of applying this result to unresolved Insufficient Information for DSE processing.

2. *Missing Not At Random conditional on both Before Followup Group and Sufficient Information Status*

This model makes the assumptions that nonresponse is nonignorable for both types of sufficiency status but the enumeration status is dependent on both the sufficiency status and Before Followup group.

3. *Missing At Random for Sufficient Information Cases/ Missing Not At Random for Insufficient Information Cases*

This model makes the assumptions that nonresponse is nonignorable for Insufficient Information for DSE processing cases but is ignorable for the Sufficient Information for DSE processing cases. Enumeration status is assumed to be dependent on both KE-status and Before Followup group.

4. *Missing At Random Conditional on Type of Response*

This model makes the assumptions that nonresponse is ignorable given the type of response. Similar to Model #1, we are also assuming that enumeration status is independent of sufficiency status, given the type of response. This can show the result of using a different covariate besides Before Followup group that is available to all cases.

5. *Missing Not At Random conditional on both Type of Response and Sufficient Information Status*

This model makes the assumptions that nonresponse is nonignorable for both types of sufficiency status but the enumeration status is dependent on both the sufficiency status and type of response.

6. *Missing At Random Using Person Followup Information and Weight Adjustments*

This model makes the assumption that the nonresponse for Insufficient Information cases is ignorable conditional on the type of response. This is implemented by a weighting adjustment using the type of response as the cells. In each cell, the weights of the unresolved Insufficient Information cases were spread to the remaining sample cases. The weights of the resolved Sufficient, resolved Insufficient and the unresolved Sufficient Information cases in the cell were adjusted upwards in the weighting adjustment.

After the weight adjustment, the nonresponse for the Sufficient Information cases is assumed to be ignorable. If the unresolved case answered the person followup question used in our analysis then we will condition on that response. If the unresolved cases did not answer the question then we will condition on the type of response for the census enumeration. This is similar to what was done for the A.C.E. Revision II missing data where they utilized followup information where available.

V. Results from the Six Example Models

This section presents results from the six models described in Section IV.C. First, we examined the results for the first three models that used Before Followup group as a covariate in the modeling. Table 3 shows the results. The table shows models that assume missingness is not at random leads to unresolved cases for most of the Before Followup groups to be imputed at rate approaching 1. This was seen when it was

assumed in Model 2 for both Sufficient and Insufficient Information cases and in Model 3 when it was only assumed for the Insufficient Information cases.

Table 3: Imputed Erroneous Enumeration Rates Assigned to Unresolved Cases for Different Models Using Before Followup Group

Before Followup Group	Sufficient Information for Matching and Followup			Insufficient Information for Matching and Followup		
	1 MAR	2 MNAR	3 MAR	1 MAR	2 MNAR	3 MNAR
Match, Followup	0.003	0.280	0.001	0.003	0.888	1.000
Match, No Followup	0.169	0.969	0.164	0.169	0[1]	0[1]
Possible Match	0.067	0.926	0.084	0.067	0.992	1.000
Nonmatch, Other Persons in Housing Unit Match	0.086	0.971	0.082	0.086	0.999	1.000
Nonmatch, Whole Household Nonmatch	0.059	0.983	0.060	0.059	0.999	1.000
Duplicate/Possible Duplicate or Fictitious/Possible Fictitious	0.998	1.000	0.998	0.998	1.000	1.000
Total	0.065	0.842	0.064	0.059	0.986	1.000

[1] The imputed result is 0 because there were no resolved Insufficient Information cases that were determined to be erroneous. All resolved cases were correct.

Using the results of these imputations combined with the sampling weights and the resolved cases, we can estimate the erroneous enumeration rate[7] using all of the cases. Table 4 shows the overall results based on the three models examined. We see that model 2 assuming Missing Not At Random for both the Sufficient and Insufficient Information cases leads to a higher report of erroneous enumerations. There is a smaller difference between Model 1 assuming Missing At Random and Model 3 assuming MNAR for Insufficient Information cases and MAR for Sufficient Information cases.

Table 4: Research Erroneous Enumeration Rate for Before Followup Group Models

	1) MAR	2) MNAR	3) MNAR/MAR
Erroneous Enumeration Rate	0.029	0.186	0.079

[7] This rate shown here is an example using the probability of the case being erroneous after missing data processing. It does not account for the probability adjustment for duplicate links within the search area used in final component estimation. See Mule (2008) Section 2.3.8 for more information.

We next examine the results from Models 4 and 5. These models use the type of response for the census enumeration assuming Missing At Random (Model 4) and Missing Not At Random (Model 5). Table 5 shows the potential benefit of this type of variable in a MAR model. For both the Sufficient and Insufficient Information cases, this MAR approach produces the same imputed erroneous rate. The imputed rate for Proxy-reported Enumerator Return (0.064) is higher than the Self-reported Enumerator return rate (0.027). The imputed rate for Self-reported mail returns is 0.008. For this MAR approach, the total row is different for the Sufficient and Insufficient Information since the weighted proportion of unresolved cases in each type of response is not the same. Since there is a higher weighted proportion of Insufficient Information unresolved cases being Non-Mailback proxy reports, this leads to an overall imputation of 0.0423. For Model 5 that uses a MNAR assumption, the result is an imputation rate of 1.00 for all of the covariates. This means that the unresolved cases are treated as full erroneous enumerations in the estimation. This MNAR result was similar to what was seen with Models 2 and 3 that used Before Followup group.

Table 5: Imputed Erroneous Enumeration Rates Based for
Two Different Assumptions Using Type of Response

Type of Response	Sufficient Information for Matching and Followup		Insufficient Information for Matching and Followup	
	4) MAR	5) MNAR	4) MAR	5) MNAR
Self-reported Mail Return	0.008	1.00	0.008	1.00
Self-reported Enumerator Return	0.027	1.00	0.027	1.00
Proxy-reported Enumerator Return	0.064	1.00	0.065	1.00
Total	0.025	1.00	0.042	1.00

Table 6 shows the overall erroneous enumeration results based on Models 4 and 5. We see that assuming Missing Not At Random for both cases again leads to a higher estimate of erroneous enumerations. The Missing Not At Random produces an estimate of 0.20 as compared to approximately 0.023 for the Missing At Random Assumption.

Table 6: Research Overall Erroneous Enumeration Rate Based on
Results of Two Different Assumption Using Types of Response Models

	4) MAR	5) MNAR
Erroneous Enumeration Rate	0.023	0.208

Finally, we examined the results from Model 6 shown in Table 7. Unresolved cases that responded that either moved or went back and forth were imputed with an erroneous enumeration rate of 0.14 as compared to 0.04 for unresolved cases that indicated they lived here all the time. This model allowed both weighting adjustments and imputation cells to be used in the missing data process. The imputed values for No PFU Response Available categories are similar to the results shown for Model 4 in Table 5. Since a

resolved case contributed as a donor to only one cell for Model 6 is the reason for the slight difference from the previous table. is In this analysis, each resolved case was only allowed to be included in one imputation cell. With imputation cell methods, resolved cases can be allowed to be included in multiple cells.

Table 7: Imputed Erroneous Enumeration Rates Based for Model 6 Using PFU Question

		Erroneous Enumeration Rate
PFU Response Available	Lived here all the time	0.041
	Moved	0.140
	Back and Forth	0.142
No PFU Response Available	Self-reported Mail Return	0.005
	Self-reported Enumerator Return	0.020
	Proxy-reported Enumerator Return	0.066
	Total	0.036

VI. Preliminary Conclusions and Future Work

This analysis has shown the application of Missing at Random and Missing Not At Random assumptions for unresolved enumeration status. The application of Missing Not At Random approaches in two instances led to most unresolved cases being imputed with a high probability of being an erroneous enumeration. This approach is based on the assumption that the nonresponse mechanism is dependent on the true enumeration status of the person. The person probably has unresolved enumeration status because of multiple other reasons for not wanting to provide their name on their questionnaire. This analysis has shown the sensitivity of the estimates to this assumption. Baker and Laird (1988) documented that the maximum likelihood estimates can result on the boundary resulting in the imputation of all erroneous as seen in our analysis.

Based on these results for Missing Not At Random assumptions, the CCM program will use a Missing at Random assumption in production for the missing data adjustment of the unresolved enumeration status for component estimation. The examination of the Before Followup group covariate shows a concern of using this covariate. Since these cells were designed for cases that went to followup, there application to the Insufficient Information cases that do not go to followup does not appear to create good imputation cells.

This examination showed some promise for the Missing At Random approach shown in Model 6. This approach used both weight adjustments and imputation cells. Since there is minimal information collected on the unresolved Insufficient Information cases, a weighting adjustment seems appropriate. By doing this adjustment, this example was able to show the ability to use PFU information in the imputation cells. This allowed unresolved cases that indicated they had moved or lived in a back and forth situation to have a higher predicted erroneous enumeration rate. The CCM will examine the 2000

A.C.E. data and the 2010 CCM questionnaires to identify appropriate weight adjustment and imputation cells that should be used. We have relied, mostly, on selecting imputation cells using an underlying knowledge of the enumeration status. Some preliminary work, not documented here, tried identifying covariates related to enumeration status using recursive partition methods. These initial partition methods invariably found covariates that ended up partitioning most of the 2006 CCM data into one, large, covariate group.

Some of the Insufficient Information for DSE processing cases were eligible to go to followup since they provided only a complete name on their Census return. Future work can examine that since these cases were eligible to go to followup that it may be more appropriate to treat these cases like the Sufficient Information for DSE processing cases.

As resources and time permit, the CCM will examine other Missing Not At Random assumptions that can be used. There are many methods introduced in the literature subsequent to the ignorable vs. nonignorable modeling approach we have employed, such as those by Stasney (1991), Nandram and Choi (2000) and Fay (1986). These newer methods make use of additional degrees of freedom in the observed data and using larger parametric models for imputation. We could look at models of the types they have proposed or others that make use of the extra information in other ways.

VII. References

Baker and Laird, (1988). "Regression analysis for categorical variables with outcome subject to nonignorable nonresponse." Journal of the American Statistical Association, v83. 62-69.

Beaghen, M. and R. Sands (2003). " Accuracy and Coverage Evaluation Revision II Missing Data Methodology and Results." *Proceedings of the Survey Research Methods Section*, American Statistical Association (2003), 485-490.

Belin, T. et al. (1993). "Hierarchical Logistic Regression Models for Imputation of Unresolved Enumeration Status in Undercount Estimation." Journal of the American Statistical Association, Vol. 88, No. 423, 1149-1159.

Cantwell, P. et al. (2001). "Missing Data Results for the Census 2000 Accuracy and Coverage Evaluation." *Proceedings of the Survey Research Methods Section*, American Statistical Association (2001).

Fay, R. (1986). "Causal Models for Nonresponse." Journal of the American Statistical Association, Vol. 81, No. 394 (1986), 354-365.

Little, R. J. and D. B. Rubin, *Statistical Analysis with Missing Data*. Wiley, second edition, 2002.

Livermore Auer, P. (2005). "Enumeration Status of Census 2000 Enumerations Deemed

Insufficient Information for Matching and Followup." *Proceedings of the Survey Research Methods Section*, American Statistical Association (2005), 2700-2707.

Moldoff, M. (2008), "The Design of the Coverage Measurement Program for the 2010 Census" DSSD 2010 Census Coverage Measurement Memorandum Series #2010-B-7.

Mule, T. (2008). "2010 Census Coverage Measurement Estimation Methodology Overview" DSSD 2010 Census Coverage Measurement Memorandum Series #2010- E-18.

Nandram, B. and J. W. Choi (2000). "Bayes Empirical Bayes Estimation of a Proportion under Nonignorable Nonresponse." *Proceedings of the Survey Research Methods Section*, American Statistical Association (2000), 215-220.

Seiss, M. and Kilmer, A. (2008). "2006 Census Coverage Measurement Procedures: Imputation Procedures and Pseudo Results of Missing Status." 2006-E-07.

Stasny, E. (1991). "Hierarchical Models for the Probabilities of a Survey Classification and Nonresponse: An Example from the National Crime Survey." Journal of the American Statistical Association, Vol. 86, 1991, 296-303.

Whitford, D. (2008). "Proposed Census Coverage Measurement Estimates for Net and Component Error." DSSD 2010 Census Coverage Measurement Memorandum Series #A-23, Washington, D.C.

Yung, W. and J. N. K. Rao (2000). "Jackknife Variance Estimation Under Imputation for Estimators Using Poststratification Information." Journal of the American Statistical Association, Vol. 95, No. 451 (Sep., 2000), 903-915.