Results of Research of Decennial Census Count Imputation Methods for Selected States

Andrew D. Kilmer
U.S. Census Bureau, Washington, DC 20233-7600

Abstract

Count imputation is the Census Bureau's process of filling in missing household population and status for addresses in the Decennial Census. In 2000, the Census Bureau used a hot deck methodology to perform count imputation. This paper describes the first phase of the Bureau's efforts to research alternative methods of count imputation for the 2010 Decennial Census. It describes the imputation methodologies being considered, describes the statistics calculated to compare the imputation methodologies, contains the results and analysis of this phase of the research, and lists the conclusions and next steps in the research.

Key Words: imputation, hot deck, spatial model, administrative records, truth deck, decennial census

1. Introduction[1]

The Census Bureau conducted count imputation for the 2000 Census to improve the accuracy of population counts and counts of occupied and vacant housing units. Count imputation had two functions: to fill in missing housing unit status (occupied, vacant, or non-existent), and to fill in missing population counts for housing units that were either known to be occupied or were imputed as occupied. There are three categories of units requiring imputation.

1. **Household Size Imputation** - The housing unit is known to be occupied, but the household size (number of household occupants) is unknown.
2. **Occupancy Imputation** - The housing unit is known to exist, but could be occupied or vacant.
3. **Status Imputation** - The address may or may not represent a valid housing unit. It could be an occupied unit, a vacant unit, or a **delete** (nonexistent unit).

For a more detailed overview of count imputation, see Chen and Kilmer (2002). Count imputation in the 2000 Census used a nearest-neighbor hot deck method. For the 2010 Census, the Census Bureau is investigating alternative count imputation methodologies in order to potentially improve count imputation. Several new methodologies are being studied. These methodologies are described in Section 2.

The Census Bureau's Imputation Work Group has created a **truth deck** to test the methodologies. The truth deck contains housing unit records from the 2000 Census that did not require count imputation. Some of these unit records were flagged to be treated as if they require imputation. The purpose of this flagging is to simulate the propensity of missing data in the 2000 Census. The flagging was replicated 100 times. For a detailed description of how the truth deck was created, see Williams (2005a and 2005b).

We tested all alternative imputation methods on the truth deck. When a given imputation methodology is executed on the truth deck, it imputes household population and status for the flagged records. The imputed values can be compared with the "true" (reported) population and status values in order to evaluate the accuracy of the methodology. Several evaluation statistics have been calculated based on these results. These statistics are detailed in Section 3. For more details on these statistics, see Chen et al. (2006).

Analysis of the imputation results was performed in two phases. The first phase was a detailed analysis of the statistics for the first three states completed[2]. This detailed analysis included calculating the state-level estimates of the evaluation statistics. It also included the following results which are not included in this paper: the variance for some statistics, estimates at the county level for some statistics, and pairwise comparisons and significance testing of some statistics. See Kilmer (2006a) for these results. The second phase of analysis, which was less detailed, involved

[1]This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed on statistical, methodological or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

[2]The states are not identified because we are evaluating the accuracy of the imputation methodologies based on their overall effect rather than their effect on any one state.

calculating only state-level estimates of the evaluation statistics for the remaining states. Analysis of the results of the first phase of analysis is in Section 4. Section 5 contains the conclusions.

2. Imputation Methodologies

This section briefly describes each of the eight count imputation methodologies that were tested on three states. For more detailed descriptions, see Appendix A of Chen et al. (2006). After the name of each methodology, the abbreviation of that methodology is in parentheses. The abbreviations will be used in various places throughout this document.

1. **2000 Hot Deck (2000 HD)**

This is the count imputation methodology used in the 2000 Census. It is a nearest neighbor hot deck method that identifies housing unit records that require imputation as “donees.” Housing unit records that do not require imputation and have data from enumerator completed forms (as a general rule) are identified as “eligible donors.” Records whose data comes from enumerator completed forms are assumed to be more similar to the donees, compared with records whose data comes from mailback forms. For a given donee, the hot deck finds a nearby eligible donor and copies the donor’s status and population to the donee. There are three imputation categories, processed in this order: household size imputation, then occupancy imputation, and finally status imputation. The methodology is described in detail in Kilmer (2002).

2. **First Modified Hot Deck (Mod-1 HD)**

The first modified hot deck is identical to the 2000 hot deck in most respects. The only difference between the two hot decks is that the first modified hot deck uses a different search algorithm. The purpose of this new algorithm is to make it more likely that the selected donor is geographically close to the donee. This algorithm is documented in detail in Kilmer (2004). That document says that there is another modification which limits (“caps”) the imputed population size at six. Since further research indicated that six may be too small a cap, no cap was used for the purpose of the research presented in this document.

3. **Second Modified Hot Deck (Mod-2 HD)**

The second modified hot deck uses the same modified search algorithm that the first modified hot deck uses. The second modified hot deck has two additional modifications. The first is different criteria for identifying an eligible donor. This criteria is meant to be a better way of identifying housing unit addresses whose data were provided by enumerators in the 2000 Census, since the hot deck donor pool is supposed to consist of housing units whose data were captured in enumerator forms. The second additional modification is processing the three imputation categories in the opposite order as the 2000 hot deck - status imputation first, occupancy imputation second, and household size imputation last. We conjectured that the order in which we processed the imputation categories in 2000 skewed the distribution of imputed status toward vacant and delete, and that reversing the order would reduce or eliminate this effect. Documentation of this hot deck is provided in Kilmer (2006b).

4. **Spatial Model 1A (SM 1a)**
5. **Spatial Model 1B (SM 1b)**
6. **Spatial Model 2 (SM 2)**

The spatial models use the relationship, at the tract[3] level, between pre-selected characteristics (called predictor variables) and household size for records not requiring imputation. Based on this relationship, the models calculate the probability of a given housing unit record having a certain population/status, and impute based on that probability. The spatial models are log linear.

The models differ from the hot decks in that the models use the characteristics of a number of housing unit records within the tract to determine the imputed population, whereas the hot decks only use one neighboring record to make this determination. Additionally, spatial modeling is a stochastic approach (meaning that the imputed value is randomly selected based on the calculated probability), while hot decks are a deterministic approach (meaning that for a given set of census responses, the hot deck will always identify the same donor for a given donee[4]). For the purposes of this research, the spatial models did not impute a household size larger than seven.

Spatial Model 1B and Spatial Model 2 are both saturated (all interactions included) models. Spatial Model 1B uses two variables to predict household size, those being mail return and nearest neighbor household type. Spatial Model 2 also uses these two variables, as well as a third variable, type of structure. Spatial Model 1A uses the same two variables as Spatial Model

[3]The tract is a level of geography defined by the Census Bureau. A tract is smaller than a county, but larger than a block.

[4]Note that there is an element of randomness in the census responses themselves. This fact is applicable to any imputation methodology we might consider.

1B but is not saturated. In the Spatial Model 1A approach, one of three models could be used for a given tract. The model uses a "forward selection" procedure which first tries a conditional independence model. If this model does not fit the data, the procedure next tries an all two-way interactions model. If this does not fit the data either, the procedure selects a saturated model (in other words, identical to Spatial Model 1B). For details on these models, see Griffin (2005 and 2006) and Sands and Griffin (2006).

7. **Administrative Records Modeling (AR Mod)**
8. **Administrative Records Direct Assignment (AR DA)**

Administrative records (AR's) are household data obtained from various government agencies. These agencies include the Census Bureau, the Internal Revenue Service, Medicare, the Social Security Administration, and others. The data from these sources were integrated into a database of housing unit and person data similar to the Census database.

Administrative records modeling uses data from both AR and census records in order to predict the status and population of a record requiring imputation. Separate models are created for each of the three imputation categories. Some examples of the predictor variables are: whether the AR address matches to a census address, whether the unit was included in non-response follow-up, whether the unit's census form was completed by an enumerator, the number of tax return person records associated with the household, and the average census household size in the unit's area. For the purposes of this research, AR modeling did not impute a household size larger than seven for status or occupancy imputation cases (but there was no limit for household size imputation cases).

Direct assignment attempts to match census units requiring imputation directly to administrative records by address. When a match is possible, the status[5] and population of the AR is used as the imputed value for the census record. For the purposes of this research, direct assignment did not impute a household size larger than seven. When a record requiring imputation could not be matched to an AR, the results of AR modeling were substituted instead. In this manner, the "direct assignment" methodology discussed in this document is actually a hybrid method of AR direct assignment and AR modeling. Both methods are documented in Farber et al. (2005 and 2006).

3. Evaluation Statistics

This section briefly describes the evaluation statistics used in this analysis. These statistics were computed at the state level for each of the eight imputation methods for the three selected states. Formulas for all statistics can be found in Chen et al. (2006).

The definitions of these statistics reference two different types of accuracy that we are attempting to measure. **Numeric accuracy** refers to how close the overall count of a particular geographic area or demographic group is to the "true" number in that area or group. **Individual accuracy** refers to the correctness of a characteristic for each individual case. The statistics discussed in this document are divided into measures of numeric accuracy and measures of individual accuracy.

3.1. Measures of Numeric Accuracy

Measures of numeric accuracy include descriptive statistics and bias.

- The **descriptive statistics** are measures of numerical accuracy of state totals. Each statistic is the relative difference between a "true" (calculated from reported data) count with what the count would have been if a given imputation methodology had been used on the given state. The computation involves all truth deck records (imputed and non-imputed) for a given state. We compute four different descriptive statistics: population, occupied units, vacant units, and deletes.
- **Bias** is a measure of how unbalanced the imputation error is, based on mutually exclusive binary outcomes, and is an indicator[6] of numerical accuracy. These binary outcomes are based on housing unit status. We compute a bias for delete/non-delete, using records flagged for status imputation, and a bias for vacant/occupied, using records flagged for occupancy imputation. The calculation involves simply

[5]In theory, it is possible for an administrative housing unit record to have a status of vacant. In practice, they are all occupied. This also implies that direct assignment cannot impute a status of vacant or delete.

[6]When we say a statistic is an "indicator" of some kind of accuracy, we mean that the statistic does not directly measure that kind of accuracy, but the result is likely related to that accuracy.

taking the difference between the number of records incorrectly imputed as the first binary outcome (delete or vacant) and the number of records incorrectly imputed as the second binary outcome (non-delete or occupied), thereby providing a measure of the bias for the given imputation category. See Thibaudeau et al. (2005) for details.

3.2. Measures of Individual Accuracy

Measures of individual accuracy include ordinal and nominal statistics and the log odds ratio.

- The **ordinal** and **nominal** statistics measure accuracy at the individual address (housing unit record) level. The computation of these statistics includes all imputed records for a given state. The ordinal statistic assigns a penalty to each error, while the nominal statistic treats all errors equally (so each imputation is either correct or incorrect). We compute ordinal and nominal statistics for population and status. The ordinal statistic for population quantifies each error based on the squared difference between the reported and imputed household sizes. The ordinal statistic for status quantifies each error as follows:
    - If the reported status is "occupied" and the imputed status is "delete," or if the reported status is "delete" and the imputed status is "occupied," the difference is defined as two.
    - Otherwise, the difference is defined as one.

  There are eight population categories: zero person (delete or vacant), one person, two persons, three persons, four persons, five persons, six persons, and seven or more persons. There are three status categories: delete, vacant, and occupied.
- The **log odds ratio** (also called the "**log crossproduct ratio**") is a measure of imputation accuracy at the address level, where "accuracy" is based on mutually exclusive binary outcomes. These binary outcomes are based on housing unit status. We compute an odds ratio for delete/non-delete using records flagged for status imputation and an odds ratio for vacant/occupied using records flagged for occupancy imputation. See Thibaudeau et al. (2005) for details.

4. Results and Analysis

This section contains the analysis of the imputation results for the three selected states (identified only as "State 1," "State 2," and "State 3"). The analysis is based on each statistic's non-parametric ranking of the eight imputation methods. We are aware of the limitations of using non-parametric rankings rather than statistical testing for this analysis, but chose the non-parametric ranking method for the sake of simplicity. Formulas for all statistics can be found in Chen et al. (2006). More detailed analysis can be found in Kilmer (2006a).

Table 1 at the end of this paper displays the results for each evaluation statistic for the three selected states. The results for a given statistic/state can be found in a given row. The methods are ranked from 1 to 8, where 1 denotes the method that obtains the best score for the given statistic and state. "Best score" is determined differently for each statistic.

- For the descriptive statistics, the best score is obtained by the method whose percent difference between total population, occupied units, vacant units, or deletes in the given state and the corresponding truth deck total is closest to zero.
- For the bias, the best score is obtained by the method whose bias is closest to zero.
- For the ordinal and nominal statistics, the best score is obtained by the method with the lowest value of the given statistic among the eight methods.
- For the log odds ratio, the best score is obtained by the method with the highest value of the given statistic among the eight methods.

Where applicable, Table 1 also indicates the direction of the error for each methodology.

- For the descriptive statistics, an "O" next to the method indicates an overcount, and a "U" indicates an undercount.
- For the bias for delete/non-delete, a "D" next to the method indicates that it is biased towards deletes, and an "N" indicates that it is biased towards non-deletes.
- For the bias for vacant/occupied, a "V" next to the method indicates that it is biased towards vacant units, and a "P" indicates that it is biased towards occupied units.

4.1. Descriptive Statistics

For the counts of population, occupied units, and vacant units, the imputation options are ordered in a similar fashion in each of the three states. Delete records are less predictable, producing very different orderings.

The spatial models are the closest to the reported population total in each state. Specifically, SM 2 is the closest in State 1 and State 3, and SM 1a is the closest in State 2. For all three states, all three spatial models are closer to the reported total than the three hot decks and the administrative data methods.

For occupied units, the eight methodologies appear in almost exactly the same order for each state. AR Modeling is the closest to the truth deck count. The three spatial models are the next best, followed by direct assignment, and finally the hot decks.

For vacant units, SM 1b is the closest to the truth deck total for State 1 and State 2, followed by the other two spatial models. In State 3, however, AR Modeling is the closest. In each state, AR Modeling and the spatial models are closer to the truth deck total than AR DA, which in turn is closer than the hot decks.

All methods generally overestimate deletes. The only exceptions are the Mod-2 hot deck and AR DA in State 3. It is difficult to say which methods most accurately estimate deletes. The second modified hot deck is closest to the reported total of deletes for State 2 and State 3, but has the fifth best total for State 1. AR direct assignment has the best score for State 1, second best for State 2, and sixth best for State 3. The spatial models fall somewhere in the middle. The 2000 hot deck and Mod-1 hot deck are worse than the other methods by this measure.

4.2. Bias

The bias results are divided into a delete/non-delete section and a vacant/occupied section.

4.2.1. Delete/Non-delete

The bias for delete records versus non-delete records is computed only for records flagged for status imputation. Records flagged for occupancy or household size imputation are excluded because they cannot be imputed with a status of delete. Note that "non-deletes" include both occupied and vacant units. For the purpose of this statistic, if a record with a reported status of occupied is imputed as vacant (or vice versa), this is considered an "accurate" imputation. Every method is generally biased in favor of deletes. This means that they are more likely to impute a reported non-delete as a delete, as opposed to imputing a reported delete as a non-delete. There are two exceptions, AR DA and Mod-2 HD, both in State 3.

No one method is clearly the least biased. The least biased method in State 1 is AR DA, while in State 2 and State 3 it is the second modified hot deck. AR Modeling is second best in State 1 and State 3. AR DA is second best in State 2. The 2000 and first modified hot decks are the most biased in each state, by a large margin. The descriptive statistics have already shown that these two methods overcount deletes, so this result is not surprising.

4.2.2. Vacant/Occupied

The bias for vacant units versus occupied units is computed only for records flagged for occupancy imputation. Records flagged for household size imputation are excluded because they cannot be imputed with a status of vacant. Records flagged for status imputation are excluded because there are three possible outcomes (occupied, vacant, or delete) and these statistics require all records to be sorted into two mutually exclusive groups.

The spatial models and administrative records methods are biased in favor of occupied units, meaning that they are more likely to impute a reported vacant unit as occupied than they are to impute a reported occupied unit as vacant. The hot decks are biased in the opposite direction. The spatial models and AR Modeling are the least biased methods in State 1. The spatial models are also the least biased methods in State 2, and AR Modeling is the least biased method in State 3. AR direct assignment is the most biased method in each state.

4.3. Ordinal and Nominal Statistics

The ordinal and nominal statistics for population rank the methods in almost exactly the same order for each state. Administrative records direct assignment outscores all other methods for both statistics and all three states. Administrative records modeling has the second best score. The third best score belongs to the second modified hot deck, followed by the other two hot decks, and lastly the spatial models.

The ordinal and nominal statistics for status rank AR direct assignment first and AR modeling second for every state, just like the ordinal and nominal statistics for population. However, unlike the population results, the ordinal and nominal statistics for status rank the spatial models ahead of the hot decks in each state. The ordinal statistic ranks the eight methods in exactly the same order as the nominal statistic in each state.

According to the descriptive statistics, AR DA was among the least accurate methods by most measures. However, the ordinal and nominal results indicate that direct assignment is the most accurate method at the housing unit level. In other words, for a given address, direct assignment is more likely to impute the correct population/status than any other method. The descriptive statistics, on the other hand, showed that direct assignment overcounts population and its population estimate is less accurate than the estimates of other methods. Using this information, we can deduce that direct assignment is biased toward imputing large household sizes. AR DA imputes the correct population more often than other methods, but when it imputes erroneously, that error is almost always in one direction - an overcount. This is especially true of addresses that have a reported status of vacant or delete, because direct assignment only imputes a status of occupied if the record matches to an administrative record. Other methods make more errors at the individual household level, but their errors could be undercounts as well as overcounts, and thus the errors cancel each other out to some degree at the state level.

The ordinal and nominal statistics for population indicate that the hot decks are more likely to impute a unit's population correctly than the spatial models. Again, this information seems to be dissonant with the descriptive statistics, which showed that the spatial models produce more accurate population counts than the hot decks. The most likely explanation is that the hot decks are more likely to err on the side of an undercount, because they have been shown to overestimate deletes and vacant units. Spatial models' errors tend to be equally likely to be overcounts and undercounts. The spatial models were not meant to be accurate at the individual address level. Instead, they were designed to produce a "smoothing" effect which is accurate at higher levels of geography, such as states or tracts.

4.4. Log Odds Ratio

The log odds ratio results are divided into a delete/non-delete section and a vacant/occupied section.

4.4.1. Delete/Non-delete

The log odds ratio for delete records versus non-delete records is computed only for records flagged for status imputation. Records flagged for occupancy or household size imputation are excluded because they cannot be imputed with a status of delete. Note that "non-deletes" include both occupied and vacant units. For the purpose of this statistic, if a record with a reported status of occupied is imputed as vacant (or vice versa), this is considered an "accurate" imputation.

For the log odds ratio, the administrative records methods score much better than the hot decks and spatial models. In this case, the results of the log odds ratio corroborate the conclusion made using the ordinal and nominal statistics, which was that the AR methodologies are more accurate at the address level than other options.

4.4.2. Vacant/Occupied

The log odds ratio for vacant units versus occupied units is computed only for records flagged for occupancy imputation. Records flagged for household size imputation are excluded because they cannot be imputed with a status of vacant. Records flagged for status imputation are excluded because there are three possible outcomes (occupied, vacant, or delete) and these statistics require all records to be sorted into two mutually exclusive groups.

The results of the log odds ratio for vacant/occupied are somewhat similar to the results of the log odds ratio for delete/non-delete. The administrative records methods score much better than the hot decks and spatial models. Direct assignment has the highest log odds ratio, and AR modeling the second highest, in all three states. The hot decks score better than the spatial models in all three states. Again, the results of the log odds ratio corroborate the conclusion made using the ordinal and nominal statistics, which was that the AR methodologies are more accurate at the address level than other options.

5. Conclusions

From this analysis, the following conclusions can be drawn:

- The second modified hot deck is generally more accurate than the 2000 hot deck or the first modified hot deck. The latter two hot decks tend to overestimate deletes by a greater margin than the other methods.
- The 2000 and first modified hot decks, by any measure, produce extremely similar results.

- The three spatial models produce similar results. In particular, Spatial Models 1A and 1B produce very similar results. In terms of numerical accuracy at the state level, these results are generally more accurate than the hot decks, but this is not necessarily true in terms of individual accuracy.
- Administrative records direct assignment overestimates population. Administrative records modeling is generally more accurate than direct assignment for population counts.
- At the housing unit level, direct assignment is the most accurate method, and administrative records modeling is the second most accurate.
- In general, all methodologies overestimate deletes.

Based on these results, the Imputation Work Group chose to exclude two imputation methods from further research: the first modified hot deck and Spatial Model 1A. The first modified hot deck was not a clear improvement over the 2000 hot deck. Spatial Model 1A produced results that were very similar to the other two spatial models, and it is more complex and uses more computer resources. The remaining six options have been run on the truth deck files for the remaining states. All statistics in this document have been computed for each state and are being analyzed.

6. References

Chen, I., and Kilmer, A. (2002). "Census 2000: Overview of Count Imputation – Reissue of Q-2." DSSD[7] Census 2000 Procedures and Operations Memorandum Series #Q-78.

Chen, I., Kilmer, A., Shores, R., and Seiss, M. (2006). "2010 Census Imputation Research: Statistics to be Computed for Count Imputation Research Options." DSSD 2006 Census Test Memorandum Series J2 (forthcoming).

Farber, J., Wagner, D., and Resnick, D. (2005). "Using Administrative Records for Imputation in the Decennial Census." Proceedings for the Section on Survey Research Methods, American Statistical Association.

Farber, J., Wagner, D., and Resnick, D. (2006). "2010 Census Count Imputation Research - Description of the Count Imputation Methods Based on Administrative Records." DSSD 2006 Census Test Memorandum Series J2 (forthcoming).

Griffin, R. (2004). "Potential Methodologies for Count Imputation for the Decennial Census." Proceedings for the Section on Survey Research Methods, American Statistical Association.

Griffin, R. (2005). "2010 Count Imputation Research - Variable Selection for Spatial Modeling." DSSD 2006 Census Test Memorandum Series #J2-06.

Griffin, R. (2006). "2010 Count Imputation Research - Documentation of Spatial Models." DSSD 2006 Census Test Memorandum Series #J2-10.

Kilmer, A. (2002). "Census 2000 Specifications for Imputing Housing Unit Status and Population Counts -- Re-Issue of Q-34." DSSD Census 2000 Procedures and Operations Memorandum Series #Q-79.

Kilmer, A. (2004). "2010 Census Count Imputation Research - Description and Effects of the Modified Hot Deck." DSSD 2006 Census Test Memorandum Series #J2-02.

Kilmer, A. (2006a). "2010 Census Count Imputation Research - Results and Analysis of Count Imputation Methodologies for Three States." DSSD 2006 Census Test Memorandum Series J2 (forthcoming).

Kilmer, A. (2006b). "2010 Census Count Imputation Research - Alternative Donor Pool and Imputation Category Processing Sequence for the Count Imputation Hot Deck." DSSD 2006 Census Test Memorandum Series #J2-08.

Sands, R., and Griffin, R. (2006). "2010 Census Count Imputation - Research Results Using Spatial Modeling." Proceedings for the Section on Survey Research Methods, American Statistical Association.

Thibaudeau, Y., Chen, I., and Sands, R. (2005). "Measuring the Discriminatory Power and Bias of Imputation Methods Designed for Imputing Status and Occupancy Status." Proceedings for the Section on Survey Research Methods, American Statistical Association.

Williams, T. (2005a). "2010 Count Imputation Research - Methodology for Developing the Truth Deck." DSSD 2006 Census Test Memorandum Series #J2-03.

Williams, T. (2005b). "The Development of Truth Decks for the 2010 Census Count Imputation Research." Proceedings for the Section on Survey Research Methods, American Statistical Association.

[7]Decennial Statistical Studies Division

Table 1: Ranking of the Eight Imputation Methodologies by Evaluation Statistics for Three States

Statistic	State	Ranking							
		1	2	3	4	5	6	7	8
Descriptive for Population	1	SM 2 U	SM 1B O	SM 1A O	AR Mod U	Mod-2 HD U	2000 HD U	Mod-1 HD U	AR DA O
	2	SM 1A U	SM 2 U	SM 1B U	AR Mod U	Mod-2 HD U	Mod-1 HD U	2000 HD U	AR DA O
	3	SM 2 U	SM 1B U	SM 1A U	Mod-2 HD U	2000 HD U	Mod-1 HD U	AR Mod U	AR DA O
Descriptive for Occupied Units	1	AR Mod O	SM 1A U	SM 1B U	SM 2 U	AR DA O	Mod-2 HD U	2000 HD U	Mod-1 HD U
	2	AR Mod O	SM 2 U	SM 1A U	SM 1B U	AR DA O	Mod-2 HD U	2000 HD U	Mod-1 HD U
	3	AR Mod U	SM 1B U	SM 1A U	SM 2 U	AR DA O	Mod-2 HD U	2000 HD U	Mod-1 HD U
Descriptive for Vacant Units	1	SM 1B U	SM 2 U	SM 1A U	AR Mod U	AR DA U	2000 HD O	Mod-1 HD O	Mod-2 HD O
	2	SM 1B U	SM 2 U	SM 1A U	AR Mod U	AR DA U	2000 HD O	Mod-1 HD O	Mod-2 HD O
	3	AR Mod U	SM 2 U	SM 1B U	SM 1A U	AR DA U	Mod-1 HD O	2000 HD O	Mod-2 HD O
Descriptive for Deletes	1	AR DA O	AR Mod O	SM 1B O	SM 1A O	Mod-2 HD O	SM 2 O	2000 HD O	Mod-1 HD O
	2	Mod-2 HD O	AR DA O	SM 2 O	SM 1A O	SM 1B O	AR Mod O	2000 HD O	Mod-1 HD O
	3	Mod-2 HD U	AR Mod O	SM 2 O	SM 1B O	SM 1A O	AR DA U	2000 HD O	Mod-1 HD O
Bias for Delete/ Nondelete	1	AR DA D	AR Mod D	SM 1B D	SM 1A D	Mod-2 HD D	SM 2 D	2000 HD D	Mod-1 HD D
	2	Mod-2 HD D	AR DA D	SM 2 D	SM 1A D	SM 1B D	AR Mod D	Mod-1 HD D	2000 HD D
	3	Mod-2 HD N	AR Mod D	AR DA N	SM 2 D	SM 1B D	SM 1A D	2000 HD D	Mod-1 HD D
Bias for Vacant/ Occupied	1	SM 1B P	SM 2 P	SM 1A P	AR Mod P	Mod-2 HD V	Mod-1 HD V	2000 HD V	AR DA P
	2	SM 2 P	SM 1B P	SM 1A P	AR Mod P	Mod-2 HD V	2000 HD V	Mod-1 HD V	AR DA P
	3	AR Mod P	SM 2 P	SM 1A P	SM 1B P	Mod-2 HD V	Mod-1 HD V	2000 HD V	AR DA P
Ordinal for Population	1	AR DA	AR Mod	Mod-2 HD	2000 HD	Mod-1 HD	SM 2	SM 1B	SM 1A
	2	AR DA	AR Mod	Mod-2 HD	Mod-1 HD	2000 HD	SM 2	SM 1A	SM 1B
	3	AR DA	AR Mod	Mod-2 HD	Mod-1 HD	2000 HD	SM 2	SM 1B	SM 1A
Nominal for Population	1	AR DA	AR Mod	Mod-2 HD	2000 HD	Mod-1 HD	SM 2	SM 1B	SM 1A
	2	AR DA	AR Mod	Mod-2 HD	Mod-1 HD	2000 HD	SM 2	SM 1B	SM 1A
	3	AR DA	AR Mod	Mod-2 HD	Mod-1 HD	2000 HD	SM 2	SM 1A	SM 1B
Ordinal for Status	1	AR DA	AR Mod	SM 1B	SM 1A	SM 2	Mod-2 HD	Mod-1 HD	2000 HD
	2	AR DA	AR Mod	SM 2	SM 1A	SM 1B	Mod-2 HD	2000 HD	Mod-1 HD
	3	AR DA	AR Mod	SM 2	SM 1B	SM 1A	Mod-2 HD	2000 HD	Mod-1 HD
Nominal for Status	1	AR DA	AR Mod	SM 1B	SM 1A	SM 2	Mod-2 HD	Mod-1 HD	2000 HD
	2	AR DA	AR Mod	SM 2	SM 1A	SM 1B	Mod-2 HD	Mod-1 HD	2000 HD
	3	AR DA	AR Mod	SM 2	SM 1B	SM 1A	Mod-2 HD	2000 HD	Mod-1 HD
Log Odds Ratio for Delete/ Nondelete	1	AR DA	AR Mod	SM 1B	SM 1A	SM 2	Mod-2 HD	2000 HD	Mod-1 HD
	2	AR DA	AR Mod	Mod-1 HD	2000 HD	Mod-2 HD	SM 2	SM 1A	SM 1B
	3	AR DA	AR Mod	2000 HD	Mod-1 HD	Mod-2 HD	SM 2	SM 1B	SM 1A
Log Odds Ratio for Vacant/ Occupied	1	AR DA	AR Mod	Mod-2 HD	Mod-1 HD	2000 HD	SM 2	SM 1B	SM 1A
	2	AR DA	AR Mod	Mod-2 HD	Mod-1 HD	2000 HD	SM 2	SM 1B	SM 1A
	3	AR DA	AR Mod	Mod-2 HD	Mod-1 HD	2000 HD	SM 2	SM 1B	SM 1A