Census 2000 Housing Unit Coverage Study

FINAL REPORT

This evaluation study reports the results of research and analysis undertaken by the U.S. Census Bureau. It is part of a broad program, the Census 2000 Testing, Experimentation, and Evaluation (TXE) Program, designed to assess Census 2000 and to inform 2010 Census planning. Findings from the Census 2000 TXE Program reports are integrated into topic reports that provide context and background for broader interpretation of results.

Diane F. Barrett, Michael Beaghen, Damon Smith, and Joseph Burcham

Decennial Statistical Studies Division and Planning, Research, and Evaluation Division

U S C E N S U S B U R E A U

Helping You Make Informed Decisions

CONTENTS

ii

LIST OF TABLES

EXECUTIVE SUMMARY

The Housing Unit Coverage Study measures the Census 2000 housing unit coverage using data from the Accuracy and Coverage Evaluation. The 2000 Accuracy and Coverage Evaluation survey was conducted in a nationwide sample of block clusters to measure the overall and differential coverage of the U.S. population and housing units. Dual system estimation was used to estimate the net coverage of housing units in Census 2000. This study also examines the estimated percentages of housing units missed as well as housing units erroneously enumerated. These two components of the dual system estimate, evaluated separately, are used to measure the completeness and accuracy of the final address list on April 1, 2000.

The Accuracy Coverage Evaluation consisted of two samples, the P-sample and the E-sample. The P-sample or population sample was an independent listing of housing units confirmed to exist in the sample block clusters on census day. The E-sample or enumeration sample was the housing units enumerated in the census in the same sample block clusters. The P-sample was matched to the E-sample. Based on this match, we calculated three coverage estimates: percent net undercount, percent P-sample nonmatches (that is percent of census misses) and percent of erroneous enumerations. This report evaluates the 2000 housing unit coverage estimates by various research categories, and where available, compares the 2000 estimates to 1990.

National Housing Unit Coverage Estimates

The overall coverage of housing units in Census 2000 was not significantly different from 1990 except for the percent of erroneous enumerations. Both censuses resulted in a net undercount of less than one percent and both missed less than four percent of the housing units. The 2000 percent (2.31 percent) of erroneous enumerations was slightly better than the 1990 percent(2.84 percent). The difference of 0.53 percentage point was statistically significant.

Net Coverage by Occupancy Status

For occupied and vacant housing units, no significant difference was observed between the 2000 and 1990 coverage.

- The net undercount for occupied housing units was 0.33 percent in 2000 and 0.53 percent in 1990.

- The net undercount for vacants was 3.37 percent in 2000, which was not

significantly different from the 4.71 percent net undercount in 1990.

- In both censuses, vacant housing units were significantly undercounted more frequently than occupied units. In Census 2000, the difference was 3.04 percentage points. In 1990, the difference between the net undercount for vacant and occupied units was 4.18 percentage points.

Coverage by Various Research Categories

Coverage by various research categories, where comparable, was for the most part consistent with 1990. However, there were areas where the housing unit coverage improved since 1990. The results that support this finding are:

- The undercount improved in the Midwest. The percent net undercount was significantly lower in 2000 (0.19 percent) than in 1990 (1.13 percent).

- Tenure had no impact on housing unit coverage. As in 1990, the 2000 coverage for renter-occupied housing units was not significantly different from that of owner-occupied housing units. The net undercount for owner-occupied housing units was 0.12 percent in 2000 and 0.37 percent in 1990. The net undercount for renter-occupied housing units was 0.57 percent in 2000 and 0.80 percent in 1990.

Note: The percent net undercount may result in a negative number, in which case it may represent an overcount. In this evaluation, we identify overcount only if the estimate is significantly different from zero.

- Housing units with Non-Hispanic Black householders (-0.45 percent net undercount not significantly different from zero) had better coverage than housing units with Non-Hispanic White and Non-Hispanic Some other race householders (0.38 percent net undercount).

- The net coverage of housing units in small multi-unit structures (2 to 9 housing units at the basic street address) was significantly better in 2000 (-0.17 percent net undercount not significantly different from zero) than in 1990 (2.25 percent net undercount).

- Occupied housing units in small multi-unit structures were overcounted (1.30 percent) in 2000, but were significantly undercounted (2.11 percent) in 1990.

- More than half (57.05 percent) of all erroneous enumerations were not housing units; that is, they were nonresidential or did not exist on census day. In 1990, "not a housing unit" (37.3 percent) and duplicates (33.4 percent) both were major reasons for erroneous enumerations. "Not a housing unit" had the highest percentage of vacant erroneous enumerations (66.0 percent) but duplicates had the highest percentage of occupied erroneous enumerations (40.7 percent). For 2000, the "not a housing unit" percentage increased to 74.29 percent while the duplicate percentage for occupied erroneous enumerations was lower at 28.69 percent.

.

Conclusion/Recommendations

Correctly enumerating vacant units continues to be a challenge for the Census Bureau. Estimates of net coverage, misses and erroneous enumerations for vacant units were significantly greater than the estimate for occupied housing units. Almost 75 percent of the vacant erroneous enumerations were attributed to the "not a housing unit" category.

Small multi-units (2 to 9 housing units at the basic street address) are still problematic for the Census Bureau. Although net coverage of housing units in small multi-units improved significantly over 1990, small multi-units had the highest percent of P-sample nonmatches and erroneous enumerations among the other sizes of structures.

Many of the Master Address File building operations in the census had high percentages of disagreement as well as high percentages of agreement with the Accuracy and Coverage Evaluation enumeration status. Block Canvassing showed a significantly higher percentage (98.9 percent) consistent for correct enumerations than any of the other census operations. This result may suggest that an operation whose primary purpose is to update the address list shows better performance.

It may be possible in future censuses to "customize" address list building operations and/or census coverage improvement operations to " target" problem situations such as small multi-unit structures with 2 to 9 housing units and vacant units to improve coverage. Clear instructions as well as training exercises for determining if vacant units meet the housing unit definition may minimize confusion of what types of units field staff should include or delete from the census address list.

1. BACKGROUND

1.1   What questions does this report answer?

The report provides answers to the following questions for Census 2000:

- What was the net coverage of housing units? How did it compare to 1990?
- Did the census misclassify vacant housing units as occupied?
- What was the coverage of housing units by various research categories (occupancy status, census region, tenure, type of structure, and so on)? How did it compare to 1990?
- What was the major reason for erroneous enumerations? How did it compare to 1990?

1.2   What was the 2000 Accuracy and Coverage Evaluation?

The 2000 Accuracy and Coverage Evaluation (A.C.E.) survey was conducted in a nationwide sample of block clusters to measure the overall and differential coverage of the U.S. population in Census 2000. As part of the survey's operations, field representatives visited each sample block cluster in the fall of 1999 and created an independent listing of housing units. The A.C.E. units were computer and/or clerically matched to the census inventory of addresses, specifically, the January 2000 version of the Decennial Master Address File or DMAF. The purpose of matching was to identify which addresses appeared on both lists and which addresses were nonmatches. Some addresses that appeared on both lists that appeared to be similar but may not be the same address were coded as possible matches. Nonmatches and possible matches were reconciled during a field followup interview. All addresses on the DMAF that were found in the sample block clusters received match codes. Similarly, all A.C.E. addresses that were found in the DMAF received match codes. Based on the match codes, a list of the A.C.E. housing units confirmed to exist within the block clusters was prepared for the person-level operations.

Person-level operations included Person Interviewing of A.C.E. households in confirmed existing housing units, a Targeted Extended Search of households whose addresses were assigned the incorrect block to update the A.C.E. inventory of housing units, and the Person Matching operation.
A.C.E. subsampling operations were also conducted during the Person phase to create the population sample, or P-sample and the enumeration sample or E-sample of housing units. The P-sample housing units were housing units listed

during the creation of the A.C.E. independent list that were confirmed to exist in the block cluster.   The E-sample housing units were a subset of housing units enumerated by the census in A.C.E. block clusters. During this time, several Census 2000 operations occurred that also added or deleted addresses from the DMAF.

To measure housing unit coverage accurately, the Final Housing Unit match was conducted between the updated A.C.E. housing unit inventory and the updated census address inventory. The stages in the Final Housing Unit Matching were similar to the initial housing unit match, which included a clerical matching and field followup of nonmatches and possible matches. There was no computer matching but there was a computer processing stage that determined which addresses required clerical matching. After final housing unit matching, each address record in the P-sample and E-sample contained a final match code. We used the final match codes to produce housing unit coverage estimates. The estimates of the percentages in each category are weighted to national totals.

1.3   What is the Housing Unit Coverage Study?

The Housing Unit Coverage Study (HUCS) is a study that measures the Census 2000 housing unit coverage using data from the Accuracy and Coverage Evaluation (A.C.E.). It uses dual system estimation to estimate a net coverage of housing units enumerated in Census 2000. The study also examines the percentage of housing units in the population sample not matched to the census (P-sample nonmatches) and the percent of housing units erroneously included in the enumeration sample (erroneous enumerations). These two components of the dual system estimate, evaluated separately, are used to measure the completeness and accuracy of the final address list used for Census 2000.

In 1990, the Census Bureau used information obtained from the Post Enumeration Survey (PES) to measure the housing unit coverage of the census. The 1990 HUCS sample consisted of half the PES sample. The Census Bureau matched the 1990 HUCS sample of independently listed addresses to the census address list. The matching results were used to produce estimates of net coverage, gross omissions, and gross erroneous enumerations of housing units (Childers 1993).

2. METHODS

2.1   Dual System Estimates (DSEs)

We use dual-system estimation to estimate the net coverage of housing units (HUs). The formula for the dual system estimate (DSE) of HUs is:

$$DSE = \frac{(C) \left(\frac{CE}{N_e}\right)}{\frac{M}{N_p}}$$

where:
C   = the count of housing units in the census (*does not include reinstated units* [1])
CE = the weighted estimate of the number of correct enumerations in the E-sample
N_e = the weighted number of E-sample housing units
M = the weighted number of P-sample matched housing units
N_p = the weighted number of P-sample housing units

2.1.1 What were reinstated units?

Before the start of the 2000 A.C.E. person matching, the census flagged housing units it thought to be potential duplicates and removed these units from the existing housing unit inventory. After the adds and deletes were identified and processed from the various census coverage improvement operations, some of these potential duplicates were **reinstated** and added to the final census housing unit inventory. Reinstated units were not in the A.C.E. universe, thus were not used in the dual system estimation. However, these reinstated housing units were included in the synthetic estimation for calculating the net undercount of housing units only. The number of reinstated housing units was not used in the calculations for estimating the percentage of housing units missed nor were they used in the calculations for estimating the percentage of erroneous enumerations. Refer to Hefter, 2001.

---

[1]   See section 2.1.1 for information about reinstated units.

2.2 Housing Unit Coverage Estimates

The following three coverage estimates were calculated to measure housing unit coverage and the completeness of the final address list on April 1, 2000:

- **Percent Net Undercount -** The net undercount ratio is defined as the ratio of the DSE minus the census count (*including reinstated units*[2]) divided by the DSE. This ratio is then multiplied by 100 for the percent net undercount.

  $$Percent\ net\ undercount = \frac{DSE - C*}{DSE} x100$$

  where:
  DSE = the formula defined in section 2.1 above
  C* = the count of housing units in the census (*includes reinstated units*[3])

  Note: The percent net undercount may result in a negative number, in which case it represents an overcount. In this evaluation, we identify overcount only if the number is significantly different from zero.

.

- **Percent P-sample Nonmatches** - Census omissions are determined by the number of housing units in the P-sample not matched to the census (P-sample nonmatches). The formula for the gross omission rate is one minus the match rate or the number of P-sample nonmatches over the number of P-sample housing units. This number is multiplied by 100 for the percent of P-sample nonmatches.

  $$Percent\ P\text{-}sample\ nonmatches = \left(1 - \frac{M}{N_p}\right) \times 100 \ \ or \ \ \frac{NM_P}{N_P} \times 100$$

  where:
  M = the weighted number of P-sample matched housing units
  NM_p = the weighted number of P-sample nonmatched housing units
  N_p = the weighted number of P-sample housing units

- **Percent Erroneous Enumeration** - Erroneous enumerations are the number of housing units that were included in the census in error. The formula for the erroneous enumeration rate is one minus the correct

---

[2] See section 2.1.1 for information about reinstated units.

[3] Ibid

enumeration rate or the number of erroneous enumerations in the E-sample over the number of E-sample housing units. This number is multiplied by 100 to obtain the percent erroneous enumerations.

$$Percent\ erroneous\ enumerations = \left(1-\frac{CE}{Ne}\right)\times100\ or\ \frac{EE}{Ne}\times100$$

where:
CE = the weighted estimate of the number of correct enumerations in the E-sample
EE = the weighted estimate of the number of erroneous enumerations in the E-sample
N_e = the weighted number of E-sample housing units

See Appendix A, Table A-1 for the weighted totals for the various components of the DSE.

Note: One may expect that the net percent undercount would be similar to the difference between the percent of P-sample nonmatches and the percent of erroneous enumerations. This is not the case. The number of reinstated housing units was used in the calculations for determining the net percent undercount but was not included in the calculations for estimating the percent of P-sample nonmatches and the percent of erroneous enumerations   The percentage of reinstated housing units have been included in Tables 2 and Tables 4 through 17 for informational purposes only.

2.3   Post-Stratification and Research Categories

The housing unit DSEs and the coverage correction factors were calculated within groupings of housing units called post-strata. The post-strata have been defined to minimize heterogeneity with respect to the housing unit census capture probability within each post-stratum (Hefter 2001). Refer to Appendix A, Table A-2 and Table A-3 which detail the occupied and vacant post-stratification housing unit groupings. The post-stratification used the following variables:

*   Occupancy Status
       Occupied
       Vacant

5

- Census Region
  Northeast
  Midwest
  South
  West

- Metropolitan Statistical Area/Type of Enumeration area (MSA/TEA)
  Large MSA Mailout/Mailback
  Medium MSA Mailout/Mailback
  Small MSA & Non-MSA Mailout /Mailback
  All Other TEAs

- Race/Hispanic Origin (determined by the race/Hispanic origin of Person 1
  listed on the Census 2000 questionnaire in the E-sample or by the reference
  person in the P-sample)
  Non-Hispanic White and Non-Hispanic Some other race
  Non-Hispanic Black
  Hispanic
  American Indian (on reservation)
  American Indian (off reservation)
  Asian

- Size of Structure (determined by the number of units at the basic street
  address-BSA)
  Single unit - one housing unit at the BSA
  Small multi-unit - 2 to 9 housing units at the BSA
  Large multi-unit -10 or more housing units at the BSA

This report examines the housing unit coverage estimates for the research
categories given above as well as one other characteristic, tenure.

- Tenure
  Owner-occupied
  Renter-occupied

2.4   Production Dual System Estimates (DSEs) versus Single Cell DSEs

Dual system estimates were calculated for 98 post-strata variables defined by the
research categories given above. The tables in this report provide net coverage
estimates using the single cell DSE. Where obtainable, the production DSEs are

also provided.

- **Production DSEs** - For some of the research categories we have obtained production dual system estimates (DSEs) by summing over appropriate post-strata. For example, the national percent net undercount estimate of the total number of housing units was calculated by adding the DSEs for all 98 post-strata used in production DSE.

  See Appendix A, Tables A-4 through A-8 for the percent net undercount, percent of P-sample nonmatches, and percent of erroneous enumerations by post-stratum for occupied and vacant units. These estimates used production DSEs.

- **Single Cell DSEs -** For those research categories that were combined in one or more post-strata (such as census region, units at basic street address, race/Hispanic origin of householder and MSA/TEA) or where the research category was not a post-stratum variable (such as tenure), we calculated the net coverage estimate using a single cell DSE within each category. This method does not take the post-stratification into account, thus the estimated undercounts may be understated.

2.5   Percent Net Undercount Comparisons to 1990

Where comparable and available, the percent net undercount, percent P-sample nonmatches, and/or the percent erroneous enumerations from the 1990 Housing Unit Coverage Study (Childers,1993) is provided in the tables for the various research categories.

2.6   Significance Testing

We used the Bonferroni multiple comparisons test to compare coverage estimates between various characteristics. Hypothesis testing was done at the 0.10 significance level.

3. LIMITS

The analysis is for the fifty states and the District of Columbia in the United States only.

The following errors are beyond the scope of this report:
- Errors in coding and matching status.
- Errors in data keying.
- Imputation errors.
- Correlation bias.

This report does not evaluate the reinstatement of duplicated housing units that were removed from the census inventory. The percentage of reinstated units has been provided in the tables for the various research categories for informational purposes only.

For Table 2 and Tables 4 through 17, *caution* should be used when comparing the difference between the percent of P-sample nonmatches and the percent of erroneous enumeration to the percent net undercount. The difference between these two components of the dual system estimator does not yield the percent net undercount one would expect. As described in Section 2. "Methods", the percent net undercount ratio was calculated using the number of units reinstated from the housing unit duplication operation, while the percent of misses and percent of erroneous enumerations calculations did not include the number of reinstated units.

The Accuracy and Coverage Evaluation Revision II did not affect the housing unit coverage results presented in this study. Only person results were revised based on the Accuracy Coverage Revision II.

4. RESULTS

4.1 What was the national net coverage of housing units in 2000? How did it compare to 1990?

The national coverage of housing units in Census 2000 resulted in a net undercount that was similar to the net undercount in 1990. Both censuses had a percent net undercount of less than 1.0 percent. Table 1 shows:

- The percent net undercount of housing units in Census 2000 was 0.61 percent, which was not significantly different from the percent net undercount in 1990 at 0.96 percent.

- For occupied housing units, no significant difference was observed between the 2000 and 1990 coverage rates. The percent net undercount was 0.33 percent in 2000 and 0.53 percent in 1990.

- The percent net undercount for vacants was 3.37 percent in 2000, which was not significantly different from the 4.71 percent net undercount in 1990.

Table 1 National Percent Net Undercount of Housing Units by Occupancy Status (standard error)

Status	2000 A.C.E.	1990 HUCS
National	0.61 (0.16)	0.96 (0.24)
Occupied	0.33 (0.13)	0.53 (0.21)
Vacant	3.37 (0.98)	4.71 (1.26)

4.2 What were the housing unit coverage estimates by occupancy status?

The net undercount for Census 2000 for both occupied and vacant units was not significantly different than in 1990. Refer to Table 2:

- In Census 2000, vacant housing units were significantly undercounted (3.37 percent) more than occupied units (0.33 percent) which resulted in a difference of 3.04 percentage points. In 1990, the difference between the

9

net undercount for vacants (4.71 percent) and occupied units (0.53 percent) was 4.18 percentage points.

Estimates of omissions and erroneous enumerations for vacant units was significantly higher than occupied housing units. Deciding whether an address identifies a housing unit is much more difficult when no one lives there. Information about vacant housing units is usually provided by proxy. The proxy respondent may not be as knowledgeable, especially about vacant boarded up units and units unfit for inhabitation. Applying the housing unit definition without the benefit of a respondent was difficult for census enumerators, thus these types of units may have been deleted from or included in the census in error.

- As in 1990, Census 2000 missed more vacant units (13.54 percent) than occupied units (2.61 percent). The differences between the percent P-sample nonmatch for vacants and the percent P-sample nonmatch for occupied units in both censuses were significant.

- Census 2000 also erroneously enumerated vacant housing units (10.50 percent) more than occupied housing units (1.51 percent). This was also the case in 1990. For both censuses, the difference between occupied and vacant erroneous enumeration estimated percents was significant

Table 2   Housing Unit Coverage Estimates by Occupancy Status (standard error)

Status	Percent P-sample Nonmatches		Percent Erroneous Enumeration		Percent Reinstated Units*	Percent Net Undercount		
	2000 A.C.E	1990 HUCS	2000 A.C.E.	1990 HUCS		2000 single cell DSE	2000 prod. DSE	1990 HUCS
Occupied	2.61 (0.11)	2.54 (0.2)	1.51 (0.07)	2.17 (0.1)	0.86	0.27 (0.13)	0.33 (0.13)	0.53 (0.21)
Vacant	13.54 (0.79)	12.67 (1.0)	10.50 (0.67)	9.24 (1.0)	1.03	2.40 (0.99)	3.37 (0.98)	4.71 (1.26)
National	**3.62 (0.15)**	**3.57 (0.2)**	**2.31 (0.11)**	**2.84 (0.20)**	**0.87**	**0.48 (0.17)**	**0.61 (0.16)**	**0.96 (0.24)**

* For informational purposes only See sections 2.1.1

4.3   Did Census 2000 misclassify vacant housing units as occupied?

Yes, there is evidence from the A.C.E. that Census 2000 misclassified vacant units as occupied. However, only a small percentage of this misclassification was attributable to unclassified imputations of persons in vacant units (not the Revised A.C.E. II. See section 3. "Limitations").

It is important to note that misclassification does not affect the overall undercoverage of housing units but more than likely had only a minimal effect on the percent net undercount of vacants as well as the net undercount of occupied housing units. While a net undercount asserts the census failed to enumerate vacant units, misclassification asserts the census enumerated the vacant unit but counted it as an occupied unit.

We used the same occupancy statuses that were used for the A.C.E. post-stratification. That is, the final status from the Hundred Percent Census Unedited File (HCUF) and the final estimation outcome code derived from the results of the A.C.E person interviewing. Further research is necessary to evaluate the extent to which the A.C.E. classification was correct.

- Among matched E-sample housing units, there were 2.84 million weighted census housing units classified as occupied that the A.C.E. classified as vacant (see Appendix A, Table A-10). The census, on the other hand, classified 1.2 million weighted units as vacant that the A.C.E. classified as occupied. Thus there was a net misclassification of about 1.6 million vacant housing units as occupied units (about 1.6 percent of the matched E-sample units).

- Of these 2.84 million units that the census classified as occupied but the A.C.E. classified as vacant, 164,359 (about six percent) had only non-data defined persons, and of these, 62,008 (0.38 percent) were attributable to unclassified imputations (see Appendix A, Table A-11). Thus, we can rule out unclassified imputation as a major source of classification error of occupancy status.

Table 3 shows the dual system estimates of the number of housing units by vacant and occupied. The undercount rates are a function of the census count and the dual system estimates (See section 2.1). The undercount of vacant units with respect to occupied units yields an estimated vacancy rate of 9.24 percent. The observed census vacancy rate was 8.99 percent.

11

Table 3 Housing Unit Dual System Estimates -Weighted Totals by Occupancy Status

	Census Count *(includes reinstated units)*	Dual System Estimate*	Percent Net Undercount
Occupied	105,463,423 91.01%	105,808,904 90.76%	0.33
Vacant	10,414,216 8.99%	10,777,553 9.24%	3.37
Total	115,877,639 100.0%	116,586,458 100.0%	0.61

*The DSE for occupied and vacant HUs does not add to the total due to rounding error.

4.4   What were the coverage estimates by census region?

Table 4 provides the Census 2000 coverage estimates for each census region. All three coverage estimates from 1990 for each census region are also given.

- The undercount in Census 2000 improved in the Midwest. The 2000 percent net undercount in this region was about one percentage point lower than the undercount in 1990. The percent undercount in 2000 was 0.19 percent versus 1.13 percent in 1990.

- The undercount in the Northeast, South and West in Census 2000 was no different than the undercount in 1990.

- In each census region, the percent P-sample nonmatches and the percent of erroneous enumeration estimates were about the same as in 1990. No statistically significant differences between the estimates were observed.

- Unlike 1990, Census 2000 missed more housing units in the Northeast (4.23 percent) than the Midwest (2.67 percent). In 1990, no statistical significant difference was observed between the percent P-sample nonmatches in the Northeast (4.1 percent) and Midwest (2.6 percent). This may be due to the higher standard errors in 1990.

- There were more housing units missed in the South (3.92 percent) than the Midwest (2.67 percent) in Census 2000 as well as in 1990 (3.9 percent versus 2.6 percent). There was no statistically significant difference

between the estimates of P-sample nonmatches for the South and Northeast in either census.

- In both Census 2000 and the 1990 census, more housing units were erroneously enumerated in the Northeast (about three percent in each census) than in the Midwest (1.8 percent and 1.2 percent).

- Housing units in the South (2.58 percent) were erroneously included in Census 2000 more often than housing units in the Midwest (1.8 percent). This was not the case in 1990. In 1990, the percent erroneous enumeration of housing units in the South (2.2 percent) was not significantly different from the Midwest (1.2 percent).

Table 4 Housing Unit Coverage Estimates by Census Region-Total (standard error)

Census Region	Percent P-sample Nonmatches		Percent Erroneous Enumeration		Percent Reinstated Units*	Percent Net Undercount	
	2000 ACE	1990 HUCS	2000 ACE	1990 HUCS		2000 single cell DSE	1990 HUCS
Northeast	4.23	4.1	2.73	3.0	1.09	0.47	0.53
	(0.34)	(0.5)	(0.21)	(0.5)		(0.40)	(0.52)
Midwest	2.67	2.6	1.80	1.2	0.69	0.19	1.13
	(0.22)	(0.4)	(0.14)	(0.2)		(0.26)	(0.43)
South	3.92	3.9	2.58	2.2	0.96	0.44	0.80
	(0.26)	(0.2)	(0.23)	(0.4)		(0.28)	(0.43)
West	3.58	2.8	2.04	1.6	0.73	0.86	1.48
	(0.35)	(0.5)	(0.19)	(0.3)		(0.40)	(0.58)

For informational purposes only See section 2.1.1.

Tables 5 and 6 show coverage estimates by census region, for occupied and vacant units. Only the percent net undercount from 1990 was available by occupancy status.

- The percent net undercount for occupied and vacant units, by census region, was similar to 1990. For each region, no significant differences were observed between the 2000 and 1990 undercounts. The lack of significance may be attributed to the relatively large standard errors.

- No one region was significantly undercounted more than any other in 2000 for occupied or vacant housing units. However, in 1990, vacant housing units in the West were significantly undercounted (9.22 percent) more than in the Midwest (1.92 percent).

- The Midwest had the lowest estimate of missed occupied housing units (1.79 percent) among the census regions. Comparisons of P-sample nonmatched rates of occupied housing units between the Midwest and each of the other regions were significantly different. No significant differences were observed for nonmatched vacant housing units among the regions.

- The percent of erroneously enumerated occupied housing units in the Midwest (1.05 percent) was significantly lower than in the Northeast (1.97 percent) and the South (1.56 percent). For vacant housing units, no evidence of significance was detected.

Table 5 Housing Unit Coverage Estimates by Census Region -Occupied Units (s.e)

Census Region	Percent P-sample Nonmatches	Percent Erroneous Enumeration	Percent Reinstated Units*	Percent Net Undercount		
				2000 single cell DSE	2000 prod DSE	1990 HUCS
Northeast	2.91 (0.25)	1.97 (0.18)	1.11	-0.14 (0.30)	na	0.30 (0.40)
Midwest	1.79 (0.15)	1.05 (0.11)	0.67	0.08 (0.18)	na	0.65 (0.38)
South	2.96 (0.20)	1.56 (0.11)	0.95	0.49 (0.22)	na	0.47 (0.30)
West	2.63 (0.29)	1.51 (0.17)	0.68	0.47 (0.33)	na	0.61 (0.55)
Total Occupied	**2.61 (0.11)**	**1.51 (0.07)**	**0.86**	**0.27 (0.13)**	**0.33 (0.13)**	**0.53 (0.21)**

* For informational purposes only See section 2.1.1   na - not available

Table 6 Housing Unit Coverage Estimates by Census Region-Vacant Units (s.e.)						
Census Region	**Percent P-sample Nonmatches**	**Percent Erroneous Enumeration**	**Percent Reinstated Units***	**Percent Net Undercount**		
				2000 single cell DSE	2000 prod DSE	1990 HUCS
Northeast	17.66 (2.01)	11.33 (1.11)	0.89	6.32 (2.85)	na	2.74 (3.69)
Midwest	12.49 (1.58)	10.3 (0.93)	0.95	1.52 (2.11)	na	1.92 (1.77)
South	12.03 (1.28)	11.16 (1.35)	1.01	-0.02 (1.42)	na	7.39 (2.96)
West	13.92 (1.75)	8.35 (0.91)	1.32	4.84 (0.91)	na	9.22 (1.91)
Total Vacant	**13.54 (0.81)**	**10.50 (0.67)**	**1.03**	**2.40 (1.39)**	**na**	**4.71 (1.26)**

* For informational purposes only See section 2.1.1   na - not available

4.5   What were the housing unit coverage estimates by tenure?

Table 7 compares the housing unit coverage estimates for owner-occupied units and renter-occupied units. The owner and renter status in the tenure variable was an important variable for the coverage of persons but not significant for the coverage of occupied housing units. Tenure was not one of the poststrata variables for producing housing unit dual system estimates. thus, the percent net undercount used the single cell dual system estimates.

- • Whether a housing unit was owner-occupied or renter-occupied had no impact on housing unit coverage. As in 1990, Census 2000 coverage for owner-occupied units was not significantly different from renter-occupied housing units. The net undercount for owner-occupied units was 0.12 percent in 2000 and 0.37 percent in 1990. The net undercount for renter-occupied housing units was 0.57 percent in 2000 and 0.80 in 1990.

- • No statistically significant differences of percent P-sample nonmatches and erroneous enumerations, by tenure, were observed. The percent of P-sample nonmatches and the percent of erroneous enumerations by tenure for 1990 was not available.

15

Table 7   Housing Unit Coverage Estimates by Tenure (standard error)

Tenure	Percent P-sample Nonmatches		Percent Erroneous Enumeration		Percent Reinstated Units*	Percent Net Undercount		
	2000 A.C.E.	1990 HUCS	2000 A.C.E.	1990 HUCS		2000 single cell DSE	2000 prod DSE	1990 HUCS
Owner-occupied	2.14 (0.11)	na	1.26 (0.07)	na	0.77	0.12 (0.13)	na	0.37 (0.21)
Renter-occupied	3.56 (0.22)	na	2.02 (0.15)	na	1.02	0.57 (0.26)	na	0.80 (0.39)

For informational purposes only See section 2.1.1   *na - not available*

4.6   What were the housing unit coverage estimates by race/Hispanic origin of householder?

We analyzed the coverage of occupied housing units, by race/Hispanic origin of the householder. The race/Hispanic origin groupings or domains were defined during person DSE processing. For housing unit DSE processing, occupied housing units were classified by the domain of the householder (person1). Refer to the Appendix, Table A-12, for the percent of E-sample housing units each domain represents. See Table 8 for the following comparisons.

Housing units with Non-Hispanic Black householders had a lower undercount than housing units with Non-Hispanic Whites and Non-Hispanic Some other race householders.

• Even though the estimated net undercount of -0.45 percent for housing units with Non-Hispanic Black householders was not significantly different from zero (that is not an overcount), it was significantly lower than the estimated net undercount of 0.38 percent for housing units with Non-Hispanic White and Non-Hispanic Some other race householders.

• However, their nonmatched percents were not significantly different. The percent of P-sample nonmatches was 2.34 percent for housing units with Non-Hispanic Black householders and 2.56 percent for housing units with Non-Hispanic Whites and Non-Hispanic Some other race householders.

• Nor were their percents of erroneous enumerations significantly different. Housing units with Non-Hispanic Black householders were erroneously

enumerated at 1.87 percent in the census, which was not significantly different than the 1.37 percent erroneous enumerations for housing units with Non-Hispanic White and Non-Hispanic Some other race householders.

Coverage was not significantly different between housing units with Hispanic householders and housing units with Non-Hispanic White and Non-Hispanic Some other race householders.

• Housing units with Hispanic householders were undercounted at 0.06 percent and housing units with Non-Hispanic Whites and Non-Hispanic Some other race householders were undercounted at 0.38 percent.

Table 8 Housing Unit Coverage Estimates by Race/Hispanic Origin of Householder in Occupied Units (standard error)

Race/Hispanic Origin of Householder	Percent P-sample Nonmatches		Percent Erroneous Enumeration		Percent Reinstated Units*	Percent Net Undercount		
	2000 A.C.E.	1990 HUCS	2000 A.C.E.	1990 HUCS		2000 single cell DSE	2000 prod DSE	1990 HUCS
Non-Hispanic White and Non-Hispanic Some other race	2.56 (0.12)	na	1.37 (0.07)	na	0.83	0.38 (0.14)	na	na
Non-Hispanic Black	2.34 (0.22)	2.8 (0.3)	1.87 (0.20)	2.1 (0.3)	0.93	-0.45 (0.29)	-0.44 (0.29)	na
Hispanic	3.01 (0.29)	3.1 (0.5)	1.98 (0.19)	2.2 (0.5)	1.00	0.06 (0.35)	0.19 (0.35)	na
Non-Hispanic Asian	3.00 (0.51)	2.2 (0.8)	2.09 (0.34)	1.3 (0.4)	0.69	0.26 (0.62)	0.22 (0.61)	na
Native Hawaiian or Pacific Islander	7.11 (2.54)	na	1.34 (0.53)	na	0.99	4.91 (2.62)	5.67 (2.82)	na
American Indian-on reservation or Alaska Native	6.64 (1.36)	na	3.79 (0.68)	na	1.22	1.78 (1.44)	1.88 (1.47)	na
American Indian -off reservation or Alaska Native	3.93 (0.95)	na	2.45 (0.44)	na	1.24	0.30 (1.00)	na	na

* For informational purposes only See section 2.1.1   na - not available

17

4.6.1   Were the comparisons of net undercount of occupied housing units by race/Hispanic origin of householder affected by tenure?

Table 9 provides the percent net undercount of housing units for three race/Hispanic origin categories by tenure.

- Tenure had no impact on the net undercount of occupied housing units by race/Hispanic origin of householder. None of the percent net undercounts between owner and renter-occupied housing units among the three race/Hispanic origin categories were significantly different.

- Although Non-Hispanic Black owner-occupied housing units as well as housing units with Hispanic owners were slightly overcounted at -0.67 and -0.70 percent, none of the percent net undercount estimates among the owner-occupied race/Hispanic origin categories were significantly different

Table 9   Percent Net Undercount by Race/Hispanic Origin and Tenure (standard error)

Tenure	Owner-occupied		Renter-occupied	
Race/Hispanic Origin	Percent Net Undercount	Percent Reinstated Units*	Percent Net Undercount	Percent Reinstated Units*
Non-Hispanic White and Some other race	0.24 (0.13)	0.74	0.78 (0.31)	1.08
Non-Hispanic Black	-0.67 (0.34)	1.0	-0.25 (0.43)	0.88
Hispanic	-0.70 (0.41)	0.98	0.66 (0.49)	1.02

For informational purposes only. See section 2.1.1.

4.7   What were the housing unit coverage estimates by size of structure?

Size (or type) of structure was based on the number of units at the basic street address (BSA) in 1990 as well as in 2000. For Census 2000, we examined coverage estimates by three size of structure categories; *single structures* with one unit at the BSA (single units), *small structures* with 2 to 9 housing units at the BSA (small multi-units), and *large structures* with 10 or more housing units at the BSA (large multi-units). In 1990, there were five type of structure categories; single, small multi-units with 2 to 9 housing units, medium multi-units with 10 to 49 housing units, large multi-units with 50 or more housing units and "other". The "other" category in 1990 included mobile homes and other unusual housing units, like boats, tents, and so forth. For Census 2000, it was difficult to identify and separate addresses of mobile homes (as well as addresses of unusual housing units). *In the 2000 estimates, the size of structure categories include mobile homes.* More than likely, mobile homes in a park are included in the large multi-unit category with 10 or more units since they shared the same basic street address (i.e. mobile home park name) and mobile homes not in a park may be in the single unit category. Refer to Table 10 for the following findings:

* For Census 2000, the net coverage among the three sizes of structures was about the same. No significant differences were observed. Note that the net coverage for housing units in small multi-unit structures (-0.17 percent) and large structures (-0.13 percent) was not significantly different from zero (that is not overcounted).

* The 2000 percent net undercount for single units was statistically the same estimate as it was in 1990 at 0.76 percent.   Note that the 2000 estimate for the single unit category includes mobile homes.

* Housing units in small multi-unit structures had a significantly higher undercount in 1990 (2.25 percent) than in 2000 (-0.17 percent). Overall coverage of small multi-unit structures in Census 2000 improved by 2.42 percentage points.

* Small multi-unit structures were the most problematic among the three sizes of structures for Census 2000. The estimated percent of P-sample nonmatches (6.94 percent) and the estimated percent of erroneous enumerations (4.78 percent) were both significantly higher than for single units (3.18 percent and 1.78 percent, respectively) and for large multi-unit structures (3.39 percent and 2.97 percent, respectively).

19

Table 10 Housing Unit Coverage Estimates by Size of Structure - Total (standard error)

Size of Structure	Percent P-sample Nonmatches	Percent Erroneous Enumeration	Percent Reinstated Units*	Percent Net Undercount		
				2000 single cell DSE	2000 prod DSE	1990 HUCS
Single Units 1 HU at BSA	3.18 (0.15)	1.78 (0.07)	0.68	0.76 (0.16)	na	0.76 (0.23)
Small Multi-units 2 to 9 HUs at BSA	6.94 (0.57)	4.78 (0.23)	2.48	-0.17 (0.64)	na	2.25 (0.65)
Large Multi-units 10 + HUs at BSA	3.39 (0.44)	2.97 (0.51)	0.57	-0.13 (0.54)	na	na [4]
National	**3.62 (0.15)**	**2.31 (0.11)**	**0.87**	**0.48 (0.17)**	**0.61 (0.16)**	**0.96 (0.24)**

* For informational purposes only See section 2.1.1   na - not available

Results are quite different when we examine coverage by occupancy status. Refer to Table 11 for estimates for occupied housing units and Table 12 for estimates of vacant units. For *vacant* units by size of structure, we used the production DSE estimates in the comparisons.

- Occupied small multi-units were overcounted at 1.30 percent in 2000 but were significantly undercounted in 1990 at 2.11 percent.

- The overcount for occupied small multi-units (1.30 percent) was significantly different from the undercount for occupied single units (0.62 percent) but not significantly different from occupied large multi-units (-0.08 percent). The estimate for large multi-units was not significantly different from zero.

- The undercount for vacant small multi-units (8.20 percent) was significantly higher than the percent net undercount for large multi-units (-0.62 percent).

- The net undercount for vacant single units in 2000 (3.31 percent) was significantly better than the undercount for vacant single units in 1990 (8.08 percent). The undercount for vacant single units improved by 4.77 percentage points.

---

[4]1990 HUCS   Types of structure.   Total
Medium multi-units (10-49 HUs)   -2.41 (1.22)
Large multi-units (50+ hus)   -0.94 (1.23)
Other (mostly mobile homes)   4.46 (1.28)

Occupancy status had no impact on the percent of housing units missed and housing units erroneously included in Census 2000 for small multi-units when compared to other sizes of structures.

- The percent of P-sample nonmatches (4.98 percent) and the percent of erroneous enumerations (3.74 percent) for occupied small multi-units were both significantly greater than for occupied single units (2.32 percent and 1.09 percent, respectively) and for large multi-units (2.39 percent and 1.89 percent, respectively).

- Vacant small multi-units (19.39 percent) were missed more often than vacant single units (13.32 percent) and vacant large multi-units (9.91 percent). The percent of erroneous enumerations for vacant small multi-units (12.93 percent) was significantly greater than vacant single units (9.77 percent) but not significantly greater than vacant large multi-units (11.08 percent). The rate of erroneously including vacant large multi-units was about the same as vacant small multi-units in Census 2000.

Table 11   Housing Unit Coverage Estimates by Size of Structure - Occupied Units (s.e.)

Size of Structure	Percent P-sample Nonmatches	Percent Erroneous Enumeration	Percent Reinstated Adds*	Percent Net Undercount		
				2000 single cell DSE	2000 prod DSE	1990 HUCS
Single Units 1 HU at BSA	2.32 (0.12)	1.09 (0.06)	0.64	0.62 (0.13)	0.63 (0.13)	0.05 (0.18)
Small Multi-units 2 to 9 HUs at BSA	4.98 (0.43)	3.74 (0.20)	2.63	-1.30 (0.48)	na	2.11 (0.59)
Large Multi-units 10+ HUs at BSA	2.39 (0.31)	1.89 (0.32)	0.60	-0.08 (0.44)	na	na[5]
Total Occupied	**2.61 (0.11)**	**1.51 (0.07)**	**0.86**	**0.27 (0.13)**	**0.33 (0.13)**	**0.53 (0.21)**

* For informational purposes only See section 2.1.1   na - not available

Table 12   Housing Unit Coverage Estimates by Size of Structure -Vacant Units (s.e)

---

[5] 1990 HUCS   Types of structure   Occupied
Medium multi-units (10-49 HUs)   -2.19 (1.12)
Large multi-units (50+ hus)   0.09 (0.52)
Other (mostly mobile homes)   4.50 (1.26)

Size of Structure	Percent P-sample Nonmatches	Percent Erroneous Enumeration	Percent Reinstated Units*	Percent Net Undercount		
				2000 single cell DSE	2000 prod DSE	1990 HUCS
Single Units 1 HU at BSA	13.32 0.88)	9.77 (0.46)	1.16	2.82 (1.12)	3.31 (1.08)	8.08 (1.54)
Small Multi-units 2 to 9 HUs at BSA	19.39 (2.54)	12.93 (0.11)	1.35	6.17 (3.45)	8.20 (3.28)	3.35 (2.43)
Large Multi-units 10 + HUs at BSA	9.91 (2.08)	11.08 (2.60)	0.33	-1.65 (2.15)	-0.62 (2.09)	na[6]
Total Vacant	**13.54**	**10.50**	**1.03**	**2.40**	**3.37 (0.98)**	**4.71 (1.26)**

* For informational purposes only See section 2.1.1   na - not available

4.7.1   How did the net undercount estimates for occupied housing units compare between owners and renters by size of structure?

Table 13 shows the percent net undercount by size of structure and tenure. From Table 13, notice that:

- The overcount for owner-occupied small multi-units (-5.95 percent net undercount) was significantly different from the coverage for owner-occupied single units (0.53 percent net undercount) and owner-occupied large multi-units (-0.60 percent net undercount). The percent net undercount for owner-occupied large multi-units was not significantly different from zero.

- Renter-occupied small multi-units (0.21 percent) were significantly undercounted more than owner-occupied small multi-units (-5.95 percent). Coverage between owner and renter occupied housing units for each of the other two sizes of structures was not significantly different.

---

[6]1990 HUCS   Types of structure.   Vacant

Medium multi-units (10-49 HUs)   -3.80 (4.28)
Large multi-units (50+ hus)   -8.19 (8.37)
Other (mostly mobile homes)   4.32 (3.81)

Table 13 Percent Net Undercount by Size of Structure and Tenure (standard error)

Tenure	Owner-occupied		Renter-occupied	
Type of Structure	Percent Net Undercount	Percent Reinstated Units*	Percent Net Undercount	Percent Reinstated Units*
Single Unit 1 HU at BSA	0.53 (0.12)	0.58	1.34 (0.33)	0.91
Small Multi-units 2 to 9 HUs at BSA	-5.95 (0.97)	5.21	0.21 (0.82)	1.83
Large Multi-units 10+ HUs at BSA	-0.60 (0.82)	0.78	0.12 (0.48)	0.56

* For informational purposes only See section 2.1.1

4.7.2 How did the net undercount estimates for occupied housing units compare by size of structure and race/Hispanic origin of the householder?

We examined the net undercount of occupied units, by size of structure, for three race/Hispanic origin categories. Refer to Table 14 for the following comparisons:

- Small multi-units with Non-Hispanic Black householders (-3.48 percent net undercount) and Non-Hispanic White and Non-Hispanic Some other race householders (-1.05 percent net undercount) were both significantly overcounted.

- Small multi-units with Non-Hispanic Black householders were overcounted more than small multi-units with Non Hispanic White and Non-Hispanic Some other race householders. The difference of 2.43 percentage points was statistically significant.

- The overcount of small multi-units with Non-Hispanic Black householders was significantly greater than single units (0.32 percent net undercount) and large multi-units (-0.12 percent net undercount) with Non-Hispanic Black householders.

- The net undercount of small multi-units with Hispanic householders (-0.70 percent) was not significantly different from the net undercount

23

of small multi-units with Non-Hispanic White or Non-Hispanic Some other race and Non-Hispanic Black householders. This lack of significance may be attributable to relatively high standard error of the estimated undercount for small multi-units with Hispanic householders.

Table 14   Percent Net Undercount by Size of Structure and Race/Hispanic Origin of the Householder (standard error)

Race/Hispanic Origin	Non-Hispanic White and Some other race		Non-Hispanic Black		Hispanic	
Size of Structure	Percent Net Undercount	Percent Reinstated Units*	Percent Net Undercount	Percent Reinstated Units*	Percent Net Undercount	Percent Reinstated Units*
Single Units 1 HU at BSA	0.62 (0.14)	0.63	0.32 (0.33)	0.65	0.79 (0.40)	0.69
Small Multi-units 2 to 9 HUs at BSA	-1.05 (0.62)	2.80	-3.48 (0.62)	2.34	-0.70 (0.98)	2.36
Large Multi-units 10+ HUs at BSA	0.23 (0.51)	0.59	-0.12 (0.86)	0.59	-0.75 (0.71)	0.63

* For informational purposes only See section 2.1.1

4.8   What were the coverage estimates by Metropolitan Statistical Area/Type of Enumeration Area (MSA/TEA) Group?

We have calculated coverage estimates, by MSA/TEA groups, which were defined during housing unit post-stratification. It appears that the size of the metropolitan statistical area had no impact on coverage in mailout/mailback areas. See Tables 15, 16, and 17 for the following results:

•   For all housing units, as well as for occupied and vacant housing units, there was no significant difference between the net undercounts for mailout/mailback areas in small, medium, or large MSAs versus all other types of enumeration areas.

24

- The net undercount for non-mailout/mailback areas shrank to almost nothing (0.22 percent) when we looked at occupied units. This may be attributed to the large percent of reinstated units in this category.

- Vacant housing units in non-mailout/mailback areas (17.90 percent) were missed more than mailout/mailback areas in medium MSAs (8.85 percent). The difference between the percent P-sample nonmatches was significant.

- In mailout/mailback areas, vacant housing units were erroneously enumerated in Census 2000 more in large MSAs (12.15 percent) than medium MSAs (7.87 percent).

Table 15 Housing Unit Coverage Estimates by Metropolitan Statistical Area/Type of Enumeration Area (MSA/TEA) Group - Total

MSA/TEA	Percent P-sample Nonmatches	Percent Erroneous Enumeration	Percent Reinstated Units*	Percent Net Undercount		
				2000 single cell DSE	2000 prod DSE	1990 HUCS
Large MSA Mailout/Mailback	3.01 (0.24)	2.13 (0.17)	0.69	0.22 (0.29)	na	na
Medium MSA Mailout/Mailback	2.41 (0.22)	1.60 (0.14)	0.42	0.41 (0.25)	na	na
Small MSA&NonMSA Mailout/Mailback	3.59 (0.34)	2.62 (0.38)	0.42	0.58 (0.35)	na	na
All Other TEAs	6.52 (0.44)	3.38 (0.14)	2.31	1.01 (0.47)	na	na

* For informational purposes only See section 2.1.1   na - not available

Table 16 Housing Unit Coverage Estimates by Metropolitan Statistical Area/Type of Enumeration Area (MSA/TEA) Group - Occupied Units (standard error)

MSA/TEA	Percent P-sample Nonmatches	Percent Erroneous Enumeration	Percent Reinstated Units*	Percent Net Undercount		
				2000 single cell DSE	2000 prod DSE	1990 HUCS
Large MSA Mailout/Mailback	2.36 (0.20)	1.57 (0.14)	0.71	0.11 (0.24)	na	na
Medium MSA Mailout/Mailback	1.85 (0.18)	1.14 (0.12)	0.43	0.30 (0.21)	na	na
Small MSA&NonMSA Mailout/Mailback	2.52 (0.21)	1.56 (0.15)	0.45	0.53 (0.26)	na	na
All Other TEAs	4.44 (0.35)	2.01 (0.11)	2.31	0.22 (0.37)	na	na
Total Occupied	**2.61 (0.11)**	**1.51 (0.07)**	**0.86**	**0.27 (0.13)**	**0.33 (0.13)**	**0.53 (0.21)**

* For informational purposes only See section 2.1.1   na - not available

Table 17 Housing Unit Coverage Rates by Metropolitan Statistical Area/Type of Enumeration Area (MSA/TEA) Group - Vacant Units (standard error)

MSA/TEA	Percent P-sample Nonmatches	Percent Erroneous Enumeration	Percent Reinstated Units*	Percent Net Undercount		
				2000 single cell DSE	2000 prod DSE	1990 HUCS
Large MSA Mailout/Mailback	12.95 (1.48)	12.15 (1.30)	0.40	0.51 (2.26)	0.82 (2.24)	na
Medium MSA Mailout/Mailback	8.85 (1.26)	7.87 (0.84)	0.22	0.84 (1.52)	0.87 (1.46)	na
Small MSA&NonMSA Mailout/Mailback	13.40 (1.86)	12.49 (2.38)	0.20	0.83 (2.12)	1.46 (1.88)	na
All Other TEAs	17.90 (1.59)	10.21 (0.59)	2.31	6.45 (2.16	6.92 (1.89)	na
Total Vacant	**13.54 (0.81)**	**10.50 (0.67)**	**1.03**	**2.40 (1.39)**	**3.37 (0.98)**	**0.53 (0.21)**

* For informational purposes only See section 2.1.1   na - not available

4.9   What were the reasons for erroneous enumerations of housing units?

The major reason for erroneous enumerations in 2000 was that the address was "not a housing unit"; that is, it was nonresidential or did not exist on Census day. We did not distinguish between those addresses that were nonresidential (that is, group quarters, commercial, uninhabitable, and so on) or nonexistent (such as vacant lots, demolished, burned down, unable to locate and so on). These have been combined into one type of erroneous enumeration category as "not a housing unit". See Table 18 for the following results:

•   More than half (57.05 percent) of all erroneous enumerations were not housing units. Of the occupied erroneous enumerations, the highest percentage (45.27 percent) was attributed to the "not a housing unit" category. The highest percentage (74.29) of vacant erroneous enumerations was also in this category.

•   In 1990, not a housing unit (37.3 percent) and duplicates (33.4 percent) both were major reasons for erroneous enumerations. Not a housing unit had the highest percentage of vacant erroneous enumerations (66.0 percent) but duplicates had the highest percentage of occupied erroneous enumerations (40.7 percent). For 2000, the not a housing unit percentage increased to 74.29 percent while the duplicate percentage for occupied erroneous enumerations was lower at 28.69 percent

.

•   Duplicates in both the 1990 and 2000 censuses accounted for a large portion of the erroneous enumerations. Even though there were more duplicates in 1990 than in 2000, the proportion of duplicates for 2000 may be understated. The percentage of duplicates did not include reinstated units. It is likely that some of the reinstated units may actually have been duplicates.

Table 18 Percentage of Erroneous Enumeration by Reason (standard error)

Reason	Total		Occupied		Vacant	
	2000 Percent	1990 Percent	2000 Percent	1990 Percent	2000 Percent	1990 Percent
Duplicates	24.81 (2.76)	33.4 (na)	28.69 (1.29)	40.7 (na)	19.13 (6.00)	17.0 (na)
Geocoding errors	16.15 (1.72)	16.2 (3.0)	23.67 (1.60)	22.3 (4.0)	5.16 (0.84)	2.8 (0.8)
Not a housing unit	57.05 (2.51)	37.3 (3.4)	45.27 (1.51)	24.4 (2.7)	74.29 (5.58)	66.0 (5.6)
Unresolved	1.99 (0.56)	2.8 (0.4)	2.37 (0.55)	2.0 (0.3)	1.42 (0.32)	4.8 (1.0)
Insufficient Information	na	10.2 (2.0)	na	10.6 (2.1)	na	9.4 (2.7)

na - not available   :

Table 19 provides the percent of E-sample by type of erroneous enumerations by occupancy status.   As noted above, the "not a housing unit" category was the major reason for erroneous enumerations in Census 2000. Vacant units erroneously enumerated as "not a housing unit" represents the highest percentage (7.80 percent) among the types of erroneous enumerations by occupancy status.

Table 19 Percent of E-Sample by Type of Erroneous Enumeration for Census 2000 (standard error)

Reason	Total	Occupied	Vacant
Duplicates	0.57 (0.80)	0.43 (0.04)	2.01 (0.70)
Geocoding errors	0.37 (0.04)	0.36 (0.04)	0.54 (0.08)
Not a housing unit	1.32 (0.06)	0.68 (0.04)	7.80 (0.43)
Unresolved	0.05 (0.01)	0.04 (0.01)	0.15 (0.03)
Total	**2.31 (0.11)**	**1.51 (0.07)**	**10.50 (0.67)**

**4.10 To what extent did the individual Master Address File (MAF) building
       operations agree with the A.C.E. enumeration status for units in the census?**

For the most part, many of the MAF building operations in the census had high percentages
of disagreement with A.C.E. erroneous enumerations but high percentages of agreement
with A.C.E correct enumerations.

The Census Bureau used addresses from a series of files and operations to update the Master
Address File (MAF). The different census operations influenced the coverage and
geocoding of addresses on the MAF. Two or more census operations may have disagreed
on whether a unit existed or not, but the final existence status in the census generally was
determined by the last operation that collected information for that unit. When two or more
operations provided disagreeing block codes for a particular address, the Census Bureau
used a scoring hierarchy to determine the official block.

For addresses that were considered erroneous enumerations by A.C.E., we examined the
level of agreement of individual MAF building operations in considering these addresses as
erroneous enumerations. We only looked at two types of erroneous enumerations, "not a
housing unit" (as described in section 4.9 above) and geocoding errors. We also examined
the level of agreement between A.C.E. and individual MAF building operations for
addresses that were considered correct enumerations by A.C.E.

The November 2000 MAF extract, which included the final census status, was compared to
the final results of the A.C.E.

For the full analysis and data tables, refer to Appendix B, "Comparisons of the MAF
Building Operations and the A.C.E. Enumeration Status".

5. CONCLUSIONS/RECOMMENDATIONS

The overall coverage of housing units in Census 2000 was not significantly different from
1990 except for the percent of erroneous enumerations. The 2000 percent of erroneous
enumerations was slightly better than the 1990 percent. The percent net undercount and the
percent of P-sample nonmatches were not significantly different between 1990 and 2000.

For some research categories, coverage of housing units improved significantly.
•   The undercount improved in the Midwest. The percent net undercount was
        significantly lower in 2000 (0.19 percent) than in 1990 (1.13 percent).

- The undercount of small multi-units (2 to 9 housing units at the basic street address) was significantly lower in 2000 (-0.17 percent) than in 1990 (2.25 percent) even though the 2000 estimates contain coverage of mobile homes.

- There was better coverage of housing units with Non-Hispanic Black householders (-0.45 percent) than with Non-Hispanic White and Non-Hispanic Some other race householders (0.38 percent).

For other research categories, coverage of housing units remain a challenge for the Census Bureau.

- Small multi-units (2 to 9 housing units at the basic street address) remain problematic for the Census Bureau. Although net coverage of housing units in small multi-units improved significantly over 1990, small multi-units had the highest percent of P-sample nonmatches (census misses) and erroneous enumerations among all sizes of structures.

- Correctly enumerating vacant units continues to be a challenge for the Census Bureau. Estimates of net coverage, misses and erroneous enumerations for vacant units was significantly greater than for occupied housing units. Almost 75 percent of the vacant erroneous enumerations were attributed to the "not a housing unit" category. Deciding whether an address identifies a housing unit is much more difficult when no one lives there. Information about vacant units is usually provided by a proxy or based on observation from the field staff. The proxy respondent (or the observation of the field staff) may not be as knowledgeable, especially about vacant boarded up units and units unfit for habitation. Confusion as to whether to include or to delete these type of vacant units from the census inventory still exists.

Many of the Master Address File building operations in the census had high percentages of disagreement as well as high percentages of agreement with the Accuracy and Coverage Evaluation enumeration status. Block Canvassing showed a significantly higher percentage (98.9 percent) consistent for correct enumerations than any of the other census operations. This result may suggest that an operation whose primary purpose is to update the address list shows better performance.

Perhaps, it may be possible in future censuses to "customize" address list building operations and/or census coverage improvement operations to " target" problem areas such as small multi-units with 2 to 9 housing units and vacant units to improve coverage. Clear directive instructions as well as training exercises for determining if vacant units meet the housing unit definition may minimize confusion of what types of units field staff should include or delete from the census address list.

6. ACKNOWLEDGMENTS

7. REFERENCES

Barrett, Diane, Beaghen, Micheal, Smith, Damon and Burcham, Joseph (2001) "*Executive Steering Committee on Accuracy and Coverage Evaluation Policy II Report Number 17:Cenus 2000 Housing Unit Coverage Study*" DSSD Census 2000 Procedures and Operations Memorandum Series #U-10R2, Bureau of the Census, October 19, 2001

Burcham, Joseph (2001) *"Accuracy and Coverage Evaluation: Creation of Input Files for Computing the Housing Unit Dual System Estimates"* DSSD Census 2000 Procedures and Operations Memorandum Series Updated Q-55, Bureau of the Census, June 26, 2001.

Childers, Danny (1992) *"The 1990 Housing Unit Coverage Study"* 1990 Decennial Census Preliminary Research and Evaluation Memorandum Series No. 193, Bureau of the Census, October 21, 1992.

Childers, Danny (1993) *"Coverage of Housing in the 1990 Decennial Census"* 1990 Decennial Census Preliminary Research and Evaluation Memorandum Series No. 253, Bureau of the Census, October 7, 1993.

Childers, Danny (2000) *"Accuracy and Coverage Evaluation: The Design Document"*, DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-1, Bureau of the Census, January 26, 2001.

Hefter, Steven P. (2002) *"Accuracy and Coverage Evaluation Survey: Computer Specifications for Housing Unit Dual System Estimation (U.S.)"* Reissue of Q-56, DSSD Census 2000 Procedures and Operations Memorandum Series Q-80, Bureau of the Census, June 25, 2002.

APPENDIX A: Additional Tables

Table A-1 Weighted Totals of Housing Units by Match Enumeration Status

DSE Variables	Weighted Totals
M = the weighted number of P-sample matched housing units	108,039,905.25
NM_p = the weighted number of P-sample nonmatched housing units	4,057,941.00
N_p = the weighted number of P-sample housing units	112,097,858.00
CE = the weighted estimate of the number of correct enumerations in the E-sample	111,781,428.00
EE = the weighted estimate of the number of erroneous enumerations in the E-sample	2,647,663.90
N_e = the weighted number of the E-sample housing units	114,429,092.03
C = the count of housing units in the census (*does not include reinstated units*)	114,858,582.00
C* = the count of housing units in the census (*includes reinstated units*)	115,877,639.00

Table A-2 Housing Unit Post-Stratification Groupings-Occupied HUs

(Definition of HUPOST- 1st digit=1, 2nd & 3rd digits given below in table)

DOMAIN	UBSA2	MSATEA	REGION			
			NE (1)	MW (2)	S (3)	W (4)
Non-Hispanic White and Non-Hispanic "Some other race" **and** Alaska Native or American Indian-off reservation	Single Units	Large MSA MO/MB (1)	01	02	03	04
		Medium MSA MO/MB (2)	05	06	07	08
		Small MSA & Non-MSA MO/MB (3)	09	10	11	12
		All Other TEAs (4)	13	14	15	16
	Small Multi-units	Large MSA MO/MB (1)	17	18	19	20
		Medium MSA MO/MB (2)	21	22	23	24
		Small MSA & Non-MSA MO/MB (3)	25	26	27	28
		All Other TEAs (4)	29	30	31	32
	Large Multi-units	Large MSA MO/MB (1)	33	34	35	36
		Medium MSA MO/MB (2)	37	38	39	40
		Small MSA & Non-MSA MO/MB (3)	41	42	43	44
		All Other TEAs (4)	45	46	47	48
Non-Hispanic Black	Single Units	Large MSA MO/MB (1)	49			
		Medium MSA MO/MB (2)	50			
		Small MSA & Non-MSA MO/MB (3)	51			
		All Other TEAs (4)	52			
	Small Multi-units	Large MSA MO/MB (1)	53			
		Medium MSA MO/MB (2)	54			
		Small MSA & Non-MSA MO/MB (3)	55			
		All Other TEAs (4)	56			
	Large Multi-units	Large MSA MO/MB (1)	57			
		Medium MSA MO/MB (2)	58			
		Small MSA & Non-MSA MO/MB (3)	59			
		All Other TEAs (4)	60			
Hispanic	Single Units	Large MSA MO/MB (1)	61			
		Medium MSA MO/MB (2)	62			
		Small MSA & Non-MSA MO/MB (3)	63			
		All Other TEAs (4)	64			
	Small Multi-units	Large MSA MO/MB (1)	65			
		Medium MSA MO/MB (2)	66			
		Small MSA & Non-MSA MO/MB (3)	67			
		All Other TEAs (4)	68			
	Large Multi-units	Large MSA MO/MB (1)	69			
		Medium MSA MO/MB (2)	70			
		Small M.A. & Non-M.A. MO/MB (3)	71			
		All Other TEAs (4)	72			

DOMAIN	UBSA2	MSATEA	ALL REGIONS
Non-Hispanic Asian	Single Units	Large MSA MO/MB (1) and Medium MSA MO/MB (2)	73
		Small MSA & Non-MSA MO/MB (3) and All Other TEAs (4)	74
	Small Multi-units	Large MSA MO/MB (1) and Medium MSA MO/MB (2)	75
		Small MSA & Non-MSA MO/MB (3) and All Other TEAs (4)	76
	Large Multi-units	Large MSA MO/MB (1) and Medium MSA MO/MB (2)	77
		Small MSA & Non-MSA MO/MB (3) and All Other TEAs (4)	78
Native Hawaiian or Pacific Islander	Single Units	Large MSA MO/MB (1) and Medium MSA MO/MB (2)	79
		Small MSA & Non-MSA MO/MB (3) and All Other TEAs (4)	80
	Small Multi-units	Large MSA MO/MB (1) and Medium MSA MO/MB (2)	81
		Small MSA & Non-MSA MO/MB (3) ans All Other TEAs (4)	82
	Large Multi-units	Large MSA MO/MB (1) and Medium MSA MO/MB (2)	83
		Small MSA & Non-MSA MO/MB (3) and All Other TEAs (4)	84
Alaska Native or American Indian - on Reservation	Single Units	85	
	Small Multi-units and Large Multi-units	86	

Table A-3 Housing Unit Post-Stratification Groupings-Vacant HUs

(Definition of HUPOST- 1st digit=0, 2nd & 3rd digits given below in table)

MSA/TEA	UBSA2		
	Single Units	**Small Multi-Units**	**Large Multi-Units**
Large MSA MO/MB	01	02	03
Medium MSA MO/MB	04	05	06
Small MSA & Non-MSA MO/MB	07	08	09
All Other TEAs	10	11	12

Table A-4	Percent Net	Undercount by Post Stratum - Occupied HUs	(s.e.) *significant			
			REGION			
DOMAIN	**UBSA2**	**MSA/TEA**	**NE** (1)	**MW** (2)	**S** (3)	**W** (4)
Non-Hispanic White and Non-Hispanic "Some other race" **and** Alaska Native or American Indian-off reservation	Single Units	Large MSA MO/MB	**0.61*** **(0.32)**	0.14 (0.28)	0.47 (0.76)	**0.66*** **(0.40)**
		Medium MSA MO/MB	1.12 (0.70)	- 0.07 (0.26)	**0.73*** **(0.37)**	0.96 (0.68)
		Small MSA & Non-MSA MO/MB	0.20 (0.77)	0.92* (0.37)	**0.93*** **(0.41)**	0.83 (0.51)
		All Other TEAs	0.12 (0.78)	- 0.46 (0.42)	**1.04*** **(0.54)**	1.30 (1.50)
	Small Multi-units	Large MSA MO/MB	0.85 (1.83)	1.33 (2.20)	- 3.80 (2.49)	**- 2.52*** **(1.20)**
		Medium MSA MO/MB	**- 3.38*** **(1.45)**	- 0.60 (1.44)	**- 4.34*** **(1.30)**	- 0.20 (2.08)
		Small MSA & Non-MSA MO/MB	4.22 (3.39)	- 0.37 (1.61)	- 0.07 (1.60)	- 1.06 (1.67)
		All Other TEAs	- 4.65 (2.88)	- 3.35 (4.13)	0.65 (8.78)	**- 13.38*** **(2.98)**
	Large Multi-units	Large MSA MO/MB	- 0.91 (0.83)	0.38 (0.68)	**4.84*** **(1.90)**	- 0.44 (1.92)
		Medium MSA MO/MB	- 6.37 (6.98)	- 2.67 (2.21)	1.34 (1.13)	0.91 (1.11)
		Small MSA & Non-MSA MO/MB	- 0.62 (0.55)	**4.38*** **(2.23)**	**- 1.39*** **(0.73)**	0.36 (5.35)
		All Other TEAs	- 1.38 (1.37)			
Non-Hispanic Black	Single Units	Large MSA MO/MB	0.12 (0.64)			
		Medium MSA MO/MB	**1.06*** **(0.47)**			
		Small MSA & Non-MSA MO/MB	- 0.05 (0.63)			
		All Other TEAs	- 0.67 (1.05)			
	Small Multi-units	Large MSA MO/MB	**- 4.20*** **(1.07)**			
		Medium MSA MO/MB	**- 2.06*** **(0.76)**			
		Small MSA & Non-MSA MO/MB	**- 3.15*** **(1.39)**			
		All Other TEAs	**- 8.77*** **(2.57)**			

Table A-4 continued

DOMAIN	UBSA2	MSA/TEA	ALL REGIONS
Non-Hispanic Black	Large Multi-units	Large MSA MO/MB	0.52 (0.74)
		Medium MSA MO/MB	0.19 (0.75)
		Small MSA & Non-MSA MO/MB	- 3.74 (5.77)
		All Other TEAs	- 0.38 (4.43)
Hispanic	Single Units	Large MSA MO/MB	0.53 (0.56)
		Medium MSA MO/MB	0.62 (0.69)
		Small MSA & Non-MSA MO/MB	**1.37 *** **(0.63)**
		All Other TEAs	1.47 (1.64)
	Small Multi-units	Large MSA MO/MB	0.50 (1.40)
		Medium MSA MO/MB	- 1.68 (1.15)
		Small MSA & Non-MSA MO/MB and All Other TEAs	- 3.39 (3.27)
	Large Multi-units	Large MSA MO/MB	- 0.62 (0.92)
		Medium MSA MO/MB	0.73 (1.10)
		Small MSA & Non-MSA MO/MB	- 3.97 (2.95)
		All Other TEAs	- 15.04 (13.25)
Non-Hispanic Asian	Single Units	Large MSA MO/MB and Medium MSA MO/MB	0.97 (0.90)
		Small MSA & Non-MSA MO/MB and All Other TEAs	1.60 (1.24)
	Small Multi-units	Large MSA MO/MB and Medium MSA MO/MB	0.87 (2.50)
		Small MSA & Non-MSA MO/MB and All Other TEAs	1.10 (5.79)
	Large Multi-units	Large MSA MO/MB and Medium MSA MO/MB	**- 1.65*** **(0.96)**
		Small MSA & Non-MSA MO/MB and All Other TEAs	- 1.18 (1.99)

DOMAIN	UBSA2	MSA/TEA	ALL REGIONS
Native Hawaiian or Pacific Islander	Single Units	Large MSA MO/MB and Medium MSA MO/MB	- 2.05 (1.52)
		Small MSA & Non-MSA MO/MB and All Other TEAs	13.41 (8.35)
	Small Multi-units	Large MSA MO/MB and Medium MSA MO/MB Small MSA & Non-MSA MO/MB and All Other TEAs	13.64 (9.20)
	Large Multi-units	Large MSA MO/MB and Medium MSA MO/MB Small MSA & Non-MSA MO/MB and All Other TEAs	2.72 (2.24)
Alaska Native or American Indian - on Reservation	Single Units	2.13 (1.53)	
	Small Multi-units and Large Multi-units (3)- 2.85	-2.85 (5.57)	

Table A-5 Percent Undercount by Post Stratum - Vacant HUs (s.e) *significant			
MSA/TEA	**UBSA2**		
	Single Units	**Small Multi-units**	**Large Multi-units**
Large MSA MO/MB	1.02 (2.76)	0.08 (4.41)	1.18 (4.31)
Medium MSA MO/MB	- 0.39 (1.96)	4.91 (4.74)	0.32 (1.70)
Small MSA & Non-MSA	2.10 (1.66)	9.78 (7.77)	- 8.10 (4.95)
All Other TEAs	**5.49*** **(1.75)**	**22.45*** **(8.92)**	8.51 (10.82)

DOMAIN	UBSA2	MSATEA	REGION			
			NE (1)	MW (2)	S (3)	W (4)
Non-Hispanic White and Non-Hispanic "Some other race" **and** Alaska Native or American Indian - off reservation	Single Units	Large MSA MO/MB	1.59 (0.30)	0.70 (0.28)	1.78 (0.62)	1.44 (0.38)
		Medium MSA MO/MB	2.31 (0.67)	0.80 (0.17)	1.69 (0.34)	1.87 (0.70)
		Small MSA & Non-MSA MO/MB	**1.91*** **(0.73)**	**2.04*** **(0.34)**	2.38 (0.39)	2.10 (0.48)
		All Other TEAs	3.18 (0.70)	2.11 (0.40)	4.84 (0.53)	5.65 (1.46)
	Small Multi-units	Large MSA MO/MB	8.73 (1.54)	5.70 (1.88)	3.24 (1.57)	2.30 (0.61)
		Medium MSA MO/MB	3.46 (1.17)	2.52 (1.40)	2.06 (0.83)	3.89 (1.93)
		Small MSA & Non-MSA MO/MB	10.77 (3.47)	3.84 (1.47)	4.14 (1.39)	3.63 (1.36)
		All Other TEAs	6.73 (2.55)	6.82 (3.53)	**13.55*** **(8.27)**	1.81 (1.09)
	Large Multi-units	Large MSA MO/MB	1.38 (0.55)	1.37 (0.61)	5.38 (1.90)	2.80 (1.16)
		Medium MSA MO/MB	**0.48*** **(0.38)**	0.44 (0.18)	3.15 (1.04)	2.15 (1.07)
		Small MSA & Non-MSA MO/MB	**0.41*** **(0.35)**	6.06 (2.15)	1.08 (0.37)	**6.31*** **(4.01)**
		All Other TEAs	2.00 (1.05)			
NonHispanic Black	Single Units	Large MSA MO/MB	1.43 (0.42)			
		Medium MSA MO/MB	1.90 (0.47)			
		Small MSA & Non-MSA MO/MB	1.99 (0.50)			
		All Other TEAs	4.79 (1.01)			
	Small Multi-units	Large MSA MO/MB	3.62 (0.78)			
		Medium MSA MO/MB	1.80 (0.49)			
		Small MSA & Non-MSA MO/MB	1.09 (0.60)			
		All Other TEAs	**2.11*** **(1.46)**			

DOMAIN	UBSA2	MSATEA	ALL REGIONS
Non-Hispanic Black	Large Multi-units	Large MSA MO/MB)	2.37 (0.69)
		Medium MSA MO/MB	1.89 (0.58)
		MSA & Non-MSA Small MO/MB	3.32 (1.48)
		All Other TEAs	**4.77*** **(3.77)**
Hispanic	Single Units	Large MSA MO/MB	1.73 (0.52)
		Medium MSA MO/MB	2.08 (0.47)
		Small MSA & Non-MSA MO/MB	2.21 (0.62)
		All Other TEAs	7.26 (1.49)
	Small Multi-units	Large MSA MO/MB	6.50 (1.31)
		Medium MSA MO/MB	2.71 (0.99)
		Small MSA & Non-MSA MO/MB and All Other TEAs	5.54 (2.69)
	Large Multi-units	Large MSA MO/MB	1.93 (0.77)
		Medium MSA MO/MB	2.34 (0.96)
		Small MSA & Non-MSA MO/MB	1.98 (0.93)
		All Other TEAs	**0.00** **(0.00)**
Non-Hispanic Asian	Single Units	Large MSA MO/MB and Medium MSA MO/MB	2.53 (0.79)
		Small MSA & Non-MSA MO/MB and All Other TEAs	3.34 (1.10)
	Small Multi-units	Large MSA MO/MB) and Medium MSA MO/MB	7.39 (2.18)
		Small MSA & Non-MSA MO/MB and All Other TEAs	9.06 (4.83)
	Large Multi-units	Large MSA MO/MB and Medium MSA MO/MB	1.24 (0.33)
		Small MSA & Non-MSA MO/MB) and All Other TEAs	2.15 (1.22)

DOMAIN	UBSA2	MSATEA	ALL REGIONS
Native Hawaiian or Pacific Islander	Single Units	Large MSA MO/MB and Medium MSA MO/MB	**0.46*** **(0.47)**
		Small MSA & Non-MSA MO/MB and All Other TEAs	15.48 (8.68)
	Small Multi-units	Large MSA MO/MB and Medium MSA MO/MB Small MSA & Non-MSA MO/MB and All Other TEAs	17.23 (9.43)
	Large Multi-units	Large MSA MO/MB and Medium MSA MO/MB Small MSA & Non-MSA MO/MB and All Other TEAs	**3.68*** **(2.24)**
Alaska Native or American Indian - on Reservation	Single Units	6.65 (1.38)	
	Small Multi-units and Large Multi-units	**6.29*** **(4.78)**	

MSATEA	UBSA2		
	Single Units	**Small Multi-Units**	**Large Multi-Units**
Large MSA MO/MB	11.31 (1.86)	19.54 (2.95)	10.51 (2.78)
Medium MSA MO/MB	9.47 (1.43)	13.24 (4.46)	5.04 (1.40)
Small MSA & Non-MSA	10.71 (1.56)	20.03 (6.75)	15.28 (6.95)
All Other TEAs	16.86 (1.50)	37.14 (7.00)	**5.64*** **(10.52)**

Table A-7 Percent P-sample Nonmatches by Post Stratum-Vacant HUs (s.e)

Table A-8 Percent Erroneous Enumerations by Post Stratum-Occupied (s.e.)*significant						
DOMAIN	**UBSA2**	**MSATEA**	**REGION**			
			NE (1)	**MW** (2)	**S** (3)	**W** (4)
Non-Hispanic White and Non-Hispanic "Some other race" and Alaska Native or American Indian - off reservation	Single Units	Large MSA MO/MB	0.81 (0.15)	0.40 (0.11)	1.13 (0.42)	0.60 (0.19)
		Medium MSA MO/MB	1.02 (0.27)	0.69 (0.24)	0.75 (0.17)	0.72 (0.24)
		Small MSA & Non-MSA MO/MB	1.49 (0.41)	0.92 (0.17)	1.23 (0.18)	1.06 (0.20)
		All Other TEAs	1.65 (0.26)	1.09 (0.15)	1.70 (0.14)	2.21 (0.32)
	Small Multi-units	Large MSA MO/MB	4.63 (0.71)	2.48 (0.66)	4.94 (1.65)	3.41 (0.88)
		Medium MSA MO/MB	3.88 (0.95)	1.55 (0.41)	4.62 (0.96)	2.51 (0.75)
		Small MSA & Non-MSA MO/MB	4.21 (1.02)	2.64 (0.62)	2.44 (0.75)	2.90 (0.92)
		All Other TEAs	3.48 (0.97)	3.60 (1.39)	6.97 (1.94)	7.74 (2.66)
	Large Multi-units	Large MSA MO/MB	1.48 (0.66)	0.56 (0.34)	**0.17*** **(0.12)**	2.82 (1.54)
		Medium MSA MO/MB	**5.95*** **(6.49)**	**2.52*** **(2.08)**	1.42 (0.54)	0.78 (0.33)
		Small MSA & Non-MSA MO/MB	**0.44*** **(0.38)**	1.19 (0.69)	1.89 (0.69)	5.38 (3.26)
		All Other TEAs	**1.27*** **(0.80)**			
Non-Hispanic Black	Single Units	Large MSA MO/MB	1.05 (0.47)			
		Medium MSA MO/MB	0.60 (0.12)			
		Small MSA & Non-MSA MO/MB	1.77 (0.49)			
		All Other TEAs	2.76 (0.48)			
	Small Multi-units	Large MSA MO/MB	4.83 (0.57)			
		Medium MSA MO/MB	2.23 (0.55)			
		Small MSA & Non-MSA MO/MB	2.82 (1.16)			
		All Other TEAs	5.82 (1.89)			

DOMAIN	UBSA2	MSATEA	ALL REGIONS
Non-Hispanic Black	Large Multi-units	Large MSA MO/MB	1.26 (0.34)
		Medium MSA MO/MB	1.23 (0.47)
		Small MSA & Non-MSA MO/MB	**6.26*** **(4.97)**
		All Other TEAs	**3.29*** **(2.12)**
Hispanic	Single Units	Large MSA MO/MB	0.96 (0.21)
		Medium MSA MO/MB	1.19 (0.50)
		Small MSA & Non-MSA MO/MB	0.57 (0.16)
		All Other TEAs	2.96 (0.57)
	Small Multi-units	Large MSA MO/MB	3.75 (0.45)
		Medium MSA MO/MB	2.52 (0.62)
		Small MSA & Non-MSA MO/MB and All Other TEAs	6.01 (1.47)
	Large Multi-units	Large MSA MO/MB	1.90 (0.48)
		Medium MSA MO/MB	1.11 (0.58)
		Small MSA & Non-MSA MO/MB	5.16 (2.63)
		All Other TEAs	**11.28*** **(10.22)**
Non-Hispanic Asian	Single Units	Large MSA MO/MB and Medium MSA MO/MB	1.34 (0.40)
		Small MSA & Non-MSA MO/MB and All Other TEAs	0.99 (0.50)
	Small Multi-units	Large MSA MO/MB and Medium MSA MO/MB	4.63 (0.98)
		Small MSA & Non-MSA MO/MB and All Other TEAs	6.02 (2.79)
	Large Multi-units	Large MSA MO/MB and Medium MSA MO/MB	2.20 (0.87)
		Small MSA & Non-MSA MO/MB and All Other TEAs	2.78 (1.62)

DOMAIN	UBSA2	MSATEA	ALL REGIONS
Native Hawaiian or Pacific Islander	Single Units	Large MSA MO/MB and Medium MSA MO/MB	**2.26*** **(1.38)**
		Small MSA & Non-MSA MO/MB and All Other TEAs	**0.62*** **(0.64)**
	Small Multi-units	Large MSA MO/MB and Medium MSA MO/MB Small MSA & Non-MSA MO/MB and All Other TEAs	2.09 (0.96)
	Large Multi-units	Large MSA MO/MB and Medium MSA MO/MB Small MSA & Non-MSA MO/MB and All Other TEAs	**0.35*** **(0.22)**
Alaska Native or American Indian - on Reservation	Single Units	3.52 (0.70)	
	Small Multi-units and Large Multi-units	6.72 (2.85)	

MSATEA	UBSA2		
	Single Units	Small Multi-units	Large Multi-units
Large MSA MO/MB	10.27* (1.81)	18.66* (2.06)	9.21* (2.82)
Medium MSA MO/MB	9.68* (1.22)	8.36* (1.68)	4.54* (1.25)
Small MSA & Non-MSA	8.70* (1.02)	10.99* (1.49)	21.39* (5.54)
All Other TEAs	10.13* (0.57)	14.76* (3.57)	6.71* (3.89)

Table A-9 Percent Erroneous Enumerations by Post Stratum-Vacant (s.e) *significant

Table A-10 Census versus A.C.E. Occupancy Status for Matched E-Sample Housing Units

Census	A.C.E.			
	Non-Interview	Occupied	Vacant	Total
Occupied	2,108,057	90,309,469	2,840,794	95,258,320
	2.06%	88.26%	2.78%	93.10%
Vacant	348,934	1,201,095	5,511,784	7,061,814
	0.34%	1.17%	5.39%	6.90%
Total	2,456,991	91,510,564	8,352,578	102,320,134
	2.40%	89.44%	8.16%	100.00%

Table A-11   Source of Final Status for the Vacant Housing Units with Non-data Defined People

Source of Final Status	Weighted Housing Units
Respondent-initiated Return	4,504
Enumerator Completed Form	97,846
Unclassified Imputation	62,008
Mail Return Checkin only	42,317
Field Status and Pop Count	4,994
Occupied Field Status but no Pop Count	14,697
Total	**164,359**

Table A-12   Percent of E-Sample by Research Category

Research Category	Characteristic	Percent
Occupancy Status	Occupied	91.06
	Vacant	8.94
Tenure	Owner-Occupied	60.76
	Renter-Occupied	30.30
Race/Hispanic Origin of Householder	Non-Hispanic White or "Some other race"	69.44
	Non-Hispanic Black	10.28
	Hispanic	7.92
	Non-Hispanic Asian	2.72
	Native Hawaiian or Pacific Islander	0.13
	Alaska Native or American Indian - on reservation	0.11
	Alaska Native or American Indian - off reservation	0.46
Size of Structure	Single	72.94
	Small multi-unit with 2-9 HUs	11.55
	Large multi-unit with 10+ HUs	15.51
	Single-Occupied	67.13
	Small multi-unit with 2-9 HUs-Occupied	10.24
	Large multi-unit with 10+ HUs-Occupied	13.68
Metropolitan Statistical Area/Type of Enumeration Area (MSA/TEA)	Large MSA MO/MB (Mailout/mailback)	28.01
	Medium MSA MO/MB	31.04
	Small MSA&NonMSA MO/MB	21.41
	All Other TEAs	19.54
	Large MSA MO/MB-Occupied	26.53
	Medium MSA MO/MB-Occupied	28.91
	Small MSA&NonMSA MO/MB-Occupied	19.33
	All Other TEAs-Occupied	16.28

APPENDIX B

To what extent do results from individual Master Address File (MAF) building operations agree with the A.C.E. enumeration status assigned to units in the census?

This section of the report analyzes our treatment by A.C.E. operations and individual MAF building operations of addresses identified as either erroneous enumerations or correct enumerations by A.C.E. The MAF building operations that we include are specifically those that were used in preparation for and/or during Census 2000.

B.1   Type of Enumeration Areas

The Census Bureau established types of enumeration areas in order to prepare for Census 2000. Each census block in the nation was classified into one of the following nine type of enumeration areas:

- Mailout/Mailback
- Update/Leave
- List/Enumerate
- Remote Alaska
- Rural Update/Enumerate
- Military
- Urban Update/Leave
- Urban Update/Enumerate
- Additions to Address Listing Universe of Blocks

The address updating methods differed in the different enumeration areas across the nation.

B.2   MAF Building Operations

The Census Bureau used addresses from a series of files and operations to update the Master Address File (MAF). The different operations influenced the coverage of addresses on the MAF and many operations influenced the geocoding of addresses on the MAF. Two or more operations may have disagreed on whether a unit existed or not, but the final existence status in the census was generally determined by the last operation that collected information for that unit. When two or more operations provided disagreeing block codes for a particular address, the Census Bureau used a scoring hierarchy to determine the official block.

The operations and address sources used to update the MAF included:

- 1990 Address Control File (ACF)
- Local Update of Census Addresses 1998 (LUCA 98)
- Block Canvassing
- LUCA 98 Field Verification
- Five Delivery Sequence Files (DSFs) from the U.S. Postal Service
- New Construction
- LUCA 98 Appeals
- Be Counted (BC)
- Telephone Questionnaire Assistance (TQA)
- Nonresponse Followup (NRFU)
- Urban Update/Enumerate
- Urban Update/Leave
- Special Place/Group Quarters Enumeration
- Coverage Improvement Followup (CIFU)
- Address Field Verification
- Census 2000 Dress Rehearsal
- Address Listing
- LUCA 99
- LUCA 99 Relisting
- LUCA 99 Appeals
- Update/Leave
- Rural Update/Enumerate
- List/Enumerate
- Remote Alaska

Each operation occurred in a unique set of enumeration areas.

B.3   Methods

All of the addresses in the analysis were considered existing housing units by the census. Results from the A.C.E. gave us information about whether a unit in the census was a correct enumeration or an erroneous enumeration. The address list building operations gave us similar information. For addresses that were considered erroneous enumerations by the A.C.E., we examined the level of agreement of individual MAF building operations in considering these addresses as erroneous enumerations. We also examined the level of agreement between the A.C.E. and individual MAF building operations for addresses that were considered correct enumerations by the A.C.E.

51

By comparing the level of agreement for erroneous enumerations to the level of agreement for correct enumerations for an individual MAF building operation, we can get a better indication of the effectiveness of the operation. For example, if an operation shows inconsistent actions for a lot of erroneous enumerations but shows consistent actions for a lot of correct enumerations, we may consider this operation to be better than an operation that shows inconsistent actions for both erroneous enumerations and correct enumerations.

Although A.C.E. results are considered the final assessment of a unit's status, it is possible for some A.C.E. results to be incorrect. In comparing MAF building operations to A.C.E., we cannot determine when a MAF building operation provides the right or wrong status for a unit. We can only determine when a MAF building operation agrees or disagrees with A.C.E. status.

The November 2000 MAF extracts were used to collect all MAF data required for this section of the evaluation. Only results for operations that contained significant numbers of addresses are present.

B.3.1 Collapsing the Types of Enumeration Areas

The nine types of enumeration areas were grouped into three categories for this evaluation:

- Inside the blue-line areas
- Outside the blue-line (Address Listing) areas, and
- Outside the blue-line (List/Enumerate) areas

The Inside the blue-line areas are composed of the Mailout/Mailback, Military, Urban Update/Leave, and Urban Update/Enumerate enumeration areas. Addresses in these areas are predominantly city-style addresses.

The Outside the blue-line (Address Listing) areas are composed of the Update/Leave, Rural Update/Enumerate, and additions to Address Listing enumeration areas. These areas contain a large number of non-city-style addresses (such as rural route addresses, P.O. Box addresses, and location descriptions).

The Outside the blue-line (List/Enumerate) areas are composed of the List/Enumerate and Remote Alaska enumeration areas. These areas also contain a large number of non-city-style addresses but these areas are more remote and

sparsely populated than the Outside the blue-line (Address Listing) areas discussed above.

We produced estimates separately for addresses inside the blue-line and addresses outside the blue-line (Address Listing) areas. Given the fact that so few operations are used to build the MAF in Outside the blue-line (List/Enumerate) areas, we did not conduct detailed analysis of addresses in these areas. We feel that any major trends will be observable in the Inside the blue-line areas and the Outside the blue-line (Address Listing) areas.

B.3.2 Types of Erroneous Enumerations

All of the addresses in our study were enumerated in the census and were later determined to be erroneous enumerations by the A.C.E. As explained earlier in the report, there were four types of erroneous enumerations in the A.C.E.:

- •   "not a housing unit" (57 percent of total erroneous enumerations)
- •   geocoding error   (16 percent)
- •   duplicate   (25 percent)
- •   unresolved type   (2 percent)

This section of the report focuses on the analysis for the first two types: "not a housing unit" and geocoding error. These two types of erroneous enumerations have a characteristic in common: for both types, the A.C.E. lister told us that the address did not exist as a housing unit in the block cluster. Therefore, we conducted the same type of analysis for both groups, and we provide the estimates separately for the two groups.

B.3.3 MAF Information Used to Analyze the Erroneous Enumerations

For a particular operation that affected both coverage and geocoding of addresses on the MAF, we have the following information:

- •   whether or not it considers a unit to be existing, and
- •   whether or not it agrees that the unit should be coded to the census block

B.3.4 The Estimates

For each MAF building operation that affected both coverage and geocoding of addresses on the MAF, eleven different estimates were produced: four for the "not a housing unit" addresses, four for geocoding errors, and three for correct enumerations..

The first estimate is the percentage of total addresses coded "not a housing unit" (or geocoding error, respectively) by the A.C.E. that the particular operation took an action on. That is, the operation made a decision about whether or not the unit existed and whether or not the unit belonged in the census block. Some operations and address sources did not have variables for both of these decisions. These operations and sources are BC, TQA, the Dress Rehearsal, and the five DSFs. The operations and sources were excluded from the analysis from the beginning.

The next three estimates are limited to units coded by the A.C.E. as "not a housing unit" (or geocoding error) and that received an action from the particular MAF building operation:

- The first of these three estimates shows the percentage of "not a housing unit" (or geocoding error) addresses that are inconsistent with the particular MAF building operation. This inconsistency is based on both the status of the unit as a housing unit and the correct geocode. The "not a housing unit" definition from the A.C.E. for these addresses does not distinguish between the two factors. The estimate includes addresses that the A.C.E. indicated were not housing units in the cluster and the MAF building operation indicated were housing units in the census block. Because the census block is in the cluster, A.C.E. also indicated these addresses were not housing units in the census block, and therefore, the A.C.E. is inconsistent with the MAF building operation.

- The second of these estimates shows the percentage of "not a housing unit" (or geocoding error) addresses that are consistent with the particular MAF building operation. The numerator for this estimate includes addresses that the A.C.E. indicated were not housing units in the cluster and the MAF building operation indicated were not housing units in the census block. Because the census block is in the cluster, both the A.C.E. and the MAF building operation indicated that the addresses were not housing units in the census block, and hence, they are consistent.

54

- The third of these estimates shows the percentage of "not a housing unit" (or geocoding error) addresses with unresolved consistency status. The numerator for this estimate includes addresses where the A.C.E. gave us information for the cluster and the MAF building operation gave us information for a block different from the census block. Because we did not conduct additional research to find out whether this other block is in the cluster or not, we cannot make any conclusions about consistency. See limits section, Block Level Analysis vs. Cluster Level Analysis subsection, for more information on the unresolved estimates.

The next estimate shows the percentage of total correct enumerations from the A.C.E. that the particular operation took an action on.

The last two estimates are limited to units coded by the A.C.E. as correct enumerations that the particular operation took an action on:

- The first of these two estimates shows the percentage of correct enumerations that are consistent with the particular MAF building operation. The numerator for this estimate includes addresses that the A.C.E. told us were housing units in the cluster and the MAF building operation told us were housing units in the census block. Because the census block is in the cluster, both the A.C.E. and the MAF building operation told us that the addresses represented housing units in the cluster. The two operations are consistent in this case.

- The last estimate shows the percentage of correct enumerations with unresolved consistency status. The numerator for this estimate includes addresses where the A.C.E. gave us information about the cluster and the MAF building operation gave us information for a block different from the census block. Because we do not know whether this other block is in the cluster or not, we cannot make any conclusions about consistency. See limits section, Block Level Analysis vs Cluster Level Analysis subsection, for more information on the unresolved estimates.

We had insufficient information to compute the percentage of correct enumerations that were inconsistent with the MAF building operation. For completeness, we include a cell for this situation in the tables, indicating "NA" for each operation.

B.4 Limits

B.4.1 *We do not know the true enumeration status when the A.C.E. disagreed with the MAF building operation*

In comparing A.C.E. results to results from the MAF building operations, we feel that A.C.E. is generally more accurate than the address list building operations in determining enumeration status. However, we do not have evidence that this is true for all addresses. Therefore, when comparing the A.C.E. to a MAF building operation, we can only refer to agreements and disagreements instead of correctness.

B.4.2 *Geographic Misallocations*

In the Methods section, we explained that we have information for a particular operation on the MAF telling us whether or not it agrees that a unit should be coded to the census block. Actually, the information tells us whether or not the operation agrees that a unit should be coded to the official block. It is possible for the official block to be different from the census block due to geographic misallocations. There were 3,026 geographic misallocation addresses that were coded as "not a housing unit" and 58 geographic misallocation addresses that were coded as geocoding errors by the A.C.E. These geographic misallocation addresses are not included in our analysis because an operation on the MAF gives us information about the blocks that these addresses were moved to and the A.C.E. gives us information about the blocks in which they originated, so we do not have the ability to compare MAF results to the A.C.E. results for these units.

B.4.3 *Block Level Analysis vs Cluster Level Analysis*

Comparing MAF results to the A.C.E. results is not as straightforward as one would like. The MAF shows us whether or not an operation codes a unit as existing in a census block while the A.C.E. shows us whether or not the unit exists in the A.C.E. sample cluster that included the census block. In many situations where the operational codes on the MAF tell us that a unit exists in a block different from the census block, we cannot make any conclusions about whether the operation is correct or incorrect without knowing if the block the unit is coded to is in the sample cluster or not. Estimates computed for these situations are known as unresolved estimates.

Another point to make is that the estimates for "not a housing unit" addresses and

geocoding errors are at the block level but the estimates for correct enumerations are at the cluster level. To illustrate this: when an operation identifies a "not a housing unit" address as existing in the census block, we know from the A.C.E. that the unit does not exist in the cluster, and therefore, does not exist in the census block (which is a block in the cluster). On the other hand, when an operation identifies a correct enumeration as existing in the block, we know from the A.C.E. that the unit does exist in the cluster, so we really don't know whether the operation places the unit in the correct block or not. This difference should be noted when comparing "not a housing unit" or geocoding error results to correct enumeration results.

B.4.4 *Suffixed Blocks*

Some blocks outside the blue-line were suffixed (divided). Each operational block flag on the MAF tells us whether or not the operation agrees that a unit belongs in the official suffixed block. Although the Address Listing operation and the LUCA 99 Relisting operation could occur in blocks that were suffixed, neither operation accounted for the suffixed block codes. These operations provided the unsuffixed block codes. Therefore, for any suffixed blocks in which Address Listing or LUCA 99 Relisting took place, the operational block flag will disagree that a unit exists in the official suffixed block, causing that unit to be in the unresolved estimate category.

This phenomenon may cause Address Listing and LUCA 99 Relisting to show lower numbers than other operations for percentage inconsistent for "not a housing unit" addresses, percentage inconsistent for geocoding errors, and percentage consistent for correct enumerations. Also, we expect these two operations to show higher numbers for the unresolved estimates.

B.4.5 *Estimates for Percentage Consistent and Percentage Inconsistent Are Not Exact*

Our estimates for percentage consistent and percentage inconsistent are not exact, because we do not know how many of the addresses in the unresolved categories should have been considered consistent or inconsistent.

B.4.6 *Lag Time*

Caution should be taken when comparing operations to each other. This is true because the operations occurred at different times, and the housing unit inventory on the ground could have changed between those times.

B.4.7 *Different Subsets of Cases*

Another reason that caution should be taken when comparing operations is that the subset of cases under consideration are different for the different operations (because only the records that the operation took an action on are considered in the calculations for each operation).

B.5   Results

This section presents estimates for each MAF building operation as described in the Methodology section. The estimates are also broken down by type of erroneous enumeration ("not a housing unit" or geocoding error) and addresses that are inside versus outside the blue-line.

The tables will present each estimate followed by the 90 percent confidence interval in parenthesis. Double asterisks in the cell of a table indicates that cell contains fewer than ten addresses in the numerator of the estimate. We do not feel that these cells contain enough addresses to provide us with reliable estimates. In the case when no addresses at all are present to compute an estimate, the cell is left blank.

Comparing the estimates "percentage of addresses that an operation took action on" indicate the impact of one operation relative to other operations on detecting erroneous enumerations. The other estimates presented measure the effectiveness of an individual operation based only on addresses that the operation took action on.

See the methods section for explanations of how to interpret the percentages in the tables.

Results presented are only for operations that took action on a significant number of addresses.

B.5.1 *Address Listing*

The Address Listing operation occurred from July 1998 through May 1999, about one year before the census enumeration. In this operation, field listers created an address list by listing all residential addresses located in "outside the blue-line (address listing)" areas.

The tables to follow below show addresses classified by the A.C.E. as "not a housing unit in the cluster," addresses classified by geocoding errors in the cluster, and addresses classified as correct enumerations in the cluster and whether or not the classifications were consistent with the classifications from the particular census operation. Table B1 presents the comparisons based on the Address Listing Operation results.

Table B1. Comparison of Address Listing and A.C.E. Classification of Addresses as "Not a Housing Unit," "Geocoding Error," and "Correct Enumeration"

A.C.E. Housing Unit Status	percent of addresses that Address Listing took action on	percent inconsistent with Address Listing	percent consistent with Address Listing	unresolved percent consistent / inconsistent
"Not a Housing Unit" total addresses = 418,367 (weighted)	67.5 (64.2,70.7)	75.1 (70.8,79.4)		24.9 (20.6,29.2)
"Geocoding Error" total addresses = 43,761 (weighted)	55.7 (45.7,65.7)	88.0 (80.7,95.2)		12.0 (4.8,19.3)
Correct Enumeration total addresses = 20,122,938 (weighted)	91.6 (90.7,92.6)	NA	77.8 (76.3,79.4)	22.2 (20.6,23.7)

Blank cells are cells that contain no addresses

In the table, we see that of all addresses enumerated in the census and later classified as "not a housing unit in the cluster" by the A.C.E., Address Listing took an action on about 68 percent of them. It also took action on 56 percent of the geocoding errors and almost 92 percent of the correct enumerations. Address Listing took an action on a relatively large percentage of the addresses because it occurred in all of the Outside the blue-line (Address Listing) areas and served as the initial listing of addresses in these areas.

As mentioned in the limits section, Address Listing did not account for suffixed block codes. Each operation on the MAF tells us whether or not it agrees that a unit belongs in the official suffixed block. Because Address Listing did not use suffixed block codes, for

any addresses in suffixed blocks, Address Listing tells us it does not agree that any of the addresses belong in the suffixed blocks. For these situations, the operation does not tell us anything about a block inside the cluster, so we cannot determine whether or not the operation agrees that the unit is in the cluster. This may explain the reason that we see relatively low numbers for percentage inconsistent for erroneous enumerations and percentage consistent for correct enumerations (compared to many other operations showing theses percentages in the 80s or 90s) and relatively high numbers for the unresolved percentages (compared to many other operations showing these percentage being less than ten) .

B.5.2 Block Canvassing

Block Canvassing occurred in the winter/spring of 1999, about one year before the census enumeration. In this operation, field listers conducted a 100 percent canvass of addresses within the "inside the blue-line" areas.

Table B2 presents comparisons between the A.C.E. and Block Canvassing classifications.

Table B2. Comparisons of Block Canvassing and A.C.E. Classification of Addresses as "Not a Housing Unit," "Geocoding Error," or "Correct Enumeration"

A.C.E. Housing Unit Status	percent of addresses that Block Canvassing took action on	percent inconsistent with Block Canvassing	percent consistent with Block Canvassing	unresolved percent consistent / inconsistent
"Not a Housing Unit" total addresses = 1,049,244 (weighted)	74.5 (71.8,77.3)	91.8 (90.4,93.1)	7.0 (5.8,8.2)	1.3 (0.8,1.7)
"Geocoding Error" total addresses = 364,101 (weighted)	82.1 (77.7,86.6)	63.4 (54.2,72.6)	6.1 (3.5,8.7)	30.5 (20.9,40.1)
Correct Enumeration total addresses = 86,621,056 (weighted)	97.8 (97.5,98.0)	NA	98.9 (98.8,99.0)	1.1 (1.0,1.2)

In this table, we see that of all addresses enumerated in the census and later classified as "not a housing unit in the cluster" by the A.C.E., Block Canvassing took an action on about 75 percent of them. Similarly, Block Canvassing took an action on 82 percent of geocoding errors and 98 percent of correct enumerations. Block Canvassing took an

action on a relatively large percentage of the addresses because it occurred in all of the areas inside the blue-line.

B.5.3 LUCA 98

In the LUCA 98 program, local governments provided updates to the MAF for addresses in the "inside the blue-line" areas. The MAF was updated with LUCA 98 addresses between October 1998 and December 1999. Refer to table B3 for comparisons to A.C.E. classification.

Table B3. Comparison of LUCA 98 and the A.C.E. in Classification of Addresses as "Not a Housing Unit," "Geocoding Error," or "Correct Enumeration"

A.C.E. Housing Unit Status	percent of addresses that LUCA 98 took action on	percent inconsistent with LUCA 98	percent consistent with LUCA 98	unresolved percent consistent / inconsistent
"Not a Housing Unit" total addresses = 1,049,244 (weighted)	16.2 (12.5,19.8)	88.2 (83.5,92.8)	4.6 (2.1,7.0)	7.3 (4.0,10.6)
"Geocoding Error" total addresses = 364,101 (weighted)	17.0 (8.5,25.5)	81.7 (60.7,100)	**	18.3 (0,39.3)
Correct Enumeration total addresses = 86,621,056 (weighted)	5.4 (4.9,5.9)	NA	88.9 (86.4,91.4)	11.1 (8.6,13.6)

** cell does not contain enough addresses in the numerator to produce a reliable estimate

Of all addresses enumerated in the census and later classified as "not a housing unit in the cluster" by the A.C.E., LUCA 98 took an action on about 16 percent of them. Similarly, LUCA 98 took an action on 17 percent of geocoding errors and 5 percent of correct enumerations. LUCA 98 occurred only in participating areas and took action on only the addresses that were provided as updates from the local governments. Therefore, it is no surprise that LUCA 98 took action on a relatively small percentage of the total addresses from the A.C.E.

B.5.4 LUCA 98 Field Verification

LUCA 98 Field Verification occurred from July 1999 through October 1999, in the "inside the blue-line" areas. After the addresses from initial phase of LUCA 98 were compared to

addresses from the Block Canvassing, the unconfirmed addresses were sent to the field for LUCA 98 Field Verification. Table B4 presents comparisons between A.C.E. and LUCA 98 Field Verification classifications.

Table B4. Comparison of LUCA 98 Field Verification and the A.C.E. Classification of Addresses as "Not a Housing Unit," "Geocoding Error," or "Correct Enumeration"

A.C.E. Housing Unit Status	percent of addresses that LUCA 98 FV took action on	percent inconsistent with LUCA 98 FV	percent consistent with LUCA 98 FV	unresolved percent consistent / inconsistent
"Not a Housing Unit"	12.5	69.7	26.0	4.4
total addresses = 1,049,244 (weighted)	(10.9,14.2)	(64.9,74.6)	(21.5,30.3)	(2.2,6.5)
"Geocoding Error"	22.5	59.4	8.7	32.0
total addresses = 364,101 (weighted)	(13.9,31.2)	(27.0,91.8)	(2.2,15.1)	(0,68.0)
Correct Enumeration	1.8	NA	85.3	14.7
total addresses = 86,621,056 (weighted)	(1.7,2.0)		(82.9,87.7)	(12.3,17.1)

We see in Table B4 that LUCA 98 Field Verification took an action on a relatively small percentage of the total addresses from the A.C.E. This is expected because the addresses sent to LUCA 98 Field Verification, in general, were the ones that were not confirmed after matching Block Canvassing results to LUCA 98 results.

We can also see from Table B4 that about 70 percent of the "not a housing unit" addresses were inconsistent and 85 percent of correct enumerations were consistent. These percentages are lower compared to that of many other operations. These relatively lower percentages may be due to the fact that LUCA 98 Field Verification could not add addresses. Instead of adding an address to the correct block, the operation only had the ability to delete an address from an incorrect block. If it would have had the opportunity to add addresses, the operation may have added more units to the census block, and therefore, would have agreed more with the final census block determination.

B.5.5 *LUCA 99 Relisting*

LUCA 99 Relisting occurred from May of 1999 through October of 1999, in the "outside the blue-line (address listing)" areas. In the Initial phase of LUCA 99, local governments reviewed block level counts of addresses from the Census Bureau. Any blocks that the local governments challenged were sent to the field for a complete relisting of the blocks. Table B5 presents comparisons between AC.E. and LUCA 99 Relisting classifications.

Table B5. Comparison of LUCA 99 Relisting and the A.C.E. Classification of Addresses as "Not a Housing Unit," "Geocoding Error," or "Correct Enumeration"

A.C.E. Housing Unit Status	percent of addresses that LUCA 99 Relisting took action on	percent inconsistent with LUCA 99 Relisting	percent consistent with LUCA 99 Relisting	unresolved percent consistent / inconsistent
"Not a Housing Unit" total addresses = 418,367 (weighted)	5.0 (3.7,6.3)	63.3 (47.8.78.8)		36.7 (21.2,52.2)
"Geocoding Error" total addresses = 43,761 (weighted)	8.9 (0,18.3)	86.1 (8.9,100)		**
Correct Enumeration total addresses = 20,122,938 (weighted)	9.7 (8.4,11.0)	NA	56.3 (48.4,64.2)	43.7 (35.8,51.6)

** cell does not contain enough addresses in the numerator to produce a reliable estimate
Blank cells are cells that contain no addresses

In Table B5, we see that of all addresses enumerated in the census and later classified as "not a housing unit in the cluster" by the A.C.E., LUCA 99 Relisting took an action on only five percent of them. Also, LUCA 99 Relisting took an action on only nine percent of the geocoding errors and ten percent of the correct enumerations. Because the only addresses relisted in LUCA 99 Relisting were in blocks that were challenged by participating governments, it makes sense that the number of addresses that LUCA 99 Relisting took an action on is small compared to the total addresses from the A.C.E.

As mentioned in the limits section and the Address Listing results section, LUCA 99 Relisting did not account for suffixed block codes. LUCA 99 Relisting tells us it does not agree that any of the addresses belong in the suffixed blocks. The operation does not tell us anything about a block inside the cluster, so we cannot determine whether or not the operation agrees that the unit is in the cluster. This may explain the reason that we see

relatively low numbers for percentage inconsistent for erroneous enumerations and percentage consistent for correct enumerations and high numbers for the unresolved percentages.

B.5.6 *Questionnaire Delivery*

Two operations were used to deliver census questionnaires to units in Outside the blue-line (Address Listing) areas. One was the Update/Leave operation, which occurred in March and April of 2000. In this operation, listers updated the address list and delivered questionnaires to the majority of the addresses in "Outside the blue-line (Address Listing)" areas.

The second operation was the Rural Update/Enumerate operation, which occurred from March of 2000 through June of 2000. In this operation, listers visited Outside the blue-line (Address Listing) areas with special enumeration needs. The listers updated the address list and at the same time enumerated the people at the units in these areas.

For the purpose of this evaluation, the two operations were put into one category called "Questionnaire Delivery." Table B6 presents comparisons between the A.C.E. and Questionnaire Delivery classifications.

Table B6. Comparison of Questionnaire Delivery and the A.C.E. Classification of Addresses as "Not a Housing Unit," "Geocoding Error," or "Correct Enumeration"

A.C.E. Housing Unit Status	percent of addresses that Questionnaire Delivery took action on	percent inconsistent with Questionnaire Delivery	percent consistent with Questionnaire Delivery	unresolved percent consistent / inconsistent
"Not a Housing Unit" total addresses = 418,367 (weighted)	65.8 (63.1,68.6)	69.0 (65.5,72.4)	31.1 (27.6,34.5)	
"Geocoding Error" total addresses = 43,761 (weighted)	50.4 (37.4,63.5)	84.0 (73.5,94.4)	10.7 (2.0,19.4)	5.3 (0.5,10.1)
Correct Enumeration total addresses = 20,122,938 (weighted)	49.0 (47.9,50.0)	NA	96.3 (96.0,96.6)	3.7 (3.4,4.0)

Blank cells are cells that contain no addresses

Because Questionnaire Delivery operations were supposed to visit all units outside the blue-line, it is a little surprising to observe in Table 6 that Questionnaire Delivery took an action on only 66 percent of addresses that were later considered "not a housing unit" by the A.C.E., 50 percent of addresses that were later considered geocoding error by the A.C.E., and 49 percent of addresses later considered correct enumerations by A.C.E. outside the blue-line. We would expect these percentages to be a lot closer to 100.

B.5.7 *Nonresponse Followup (NRFU)*

NRFU occurred from April 2000 through June 2000, immediately after the initial phase of the census enumeration. Any addresses for which the Census Bureau did not receive a completed questionnaire by April 11, 2000 were sent to NRFU to be enumerated in the field. NRFU occurred both inside the blue-line and outside the blue-line (Address Listing) areas. Table B7 presents results for addresses inside the blue-line and Table B8 presents results for addresses outside the blue-line (Address Listing) areas.

Table B7. Comparison of NRFU and the A.C.E. Classification of Addresses as "Not a Housing Unit," "Geocoding Error," or "Correct Enumeration "Inside the Blue-Line

A.C.E. Housing Unit Status	percent of addresses that NRFU took action on	percent inconsistent with NRFU	percent consistent with NRFU	unresolved percent consistent / inconsistent
"Not a Housing Unit" total addresses = 1,049,244 (weighted)	71.2 (69.1,73.4)	84.4 (80.3,88.5)	15.4 (11.3,19.5)	**
"Geocoding Error" total addresses = 364,101(weighted)	40.8 (36.8,44.8)	86.1 (81.7,90.5)	5.1 (2.0,8.3)	8.8 (5.6,12.0)
Correct Enumeration total addresses = 86,621,056 (weighted)	30.4 (30.0,30.7)	NA	97.6 (97.4,97.8)	2.4 (2.2,2.6)

** cell does not contain enough addresses in the numerator to produce a reliable estimate

Of the addresses inside the blue-line, NRFU took an action on a relatively high percentage of addresses that were determined to be "not a housing unit" by the A.C.E. and it took an action on a relatively low percentage of addresses that were determined to be geocoding errors or correct enumerations by the A.C.E. Because NRFU was large and was one of the last operations in the census, it makes sense that a high percentage of addresses considered to be "not a housing unit" from the A.C.E. were enumerated by NRFU. Similar results are shown in the Outside the blue-line (Address Listing) areas, except that the percent for geocoding errors is about the same as "not a housing unit."

Table B8. Comparison of NRFU and the A.C.E. Classification of Addresses as "Not a Housing Unit," "Geocoding Error," or "Correct Enumeration" Outside the Blue-Line (Address Listing) areas

A.C.E. Housing Unit Status	percent of addresses that NRFU took action on	percent inconsistent with NRFU	percent consistent with NRFU	unresolved percent consistent / inconsistent
"Not a Housing Unit" total addresses = 418,367 (weighted)	64.4 (61.3,67.6)	82.4 (79.7,85.2)	17.3 (14.5,20.0)	**
"Geocoding Error" total addresses = 43,761 (weighted)	60.1 (50.7,69.5)	79.0 (63.6,94.3)	1.0 (0.03,2.0)	20.0 (4.6,35.4)
Correct Enumeration total addresses = 20,122,938 (weighted)	35.4 (34.6,36.1)	NA	96.6 (96.3,96.9)	3.4 (3.1,3.7)

**cell does not contain enough addresses in the numerator to produce a reliable estimate

B.5.8 *Coverage Improvement Followup (CIFU)*

CIFU occurred in the summer/fall of 2000, immediately after NRFU. The primary purpose of CIFU was to improve coverage of housing units that may have been inaccurately classified as vacant or nonexistent in an earlier census operation. CIFU occurred both inside the blue-line and in the Outside the blue-line (Address Listing) areas. Table B9 presents results for addresses inside the blue-line and Table B10 presents results for addresses in the Outside the blue-line (Address Listing) areas.

Table B9. Comparison of CIFU and the A.C.E. Classification of addresses as "Not a Housing Unit," "Geocoding Error," or "Correct Enumeration" Inside the Blue-Line

A.C.E. Housing Unit Status	percent of addresses that CIFU took action on	percent inconsistent with CIFU	percent consistent with CIFU	unresolved percent consistent / inconsistent
"Not a Housing Unit" total addresses = 1,049,244 (weighted)	25.3 (22.7,27.9)	80.1 (76.5,83.8)	19.6 (16.0,23.2)	**
"Geocoding Error" total addresses = 364,101(weighted)	13.4 (10.0,16.7)	93.0 (89.0,97.0)	7.0 (3.0,11.0)	**
Correct Enumeration total addresses = 86,621,056 (weighted)	4.0 (3.8,4.2)	NA	95.9 (95.5,96.3)	4.1 (3.7,4.5)

** cell does not contain enough addresses in the numerator to produce a reliable estimate

In Table B9, we see that, of all addresses enumerated in the census inside the blue-line and later classified as "not a housing unit in the cluster" by the A.C.E., CIFU took an action on about 25 percent of them. Similarly, CIFU took an action on 13 percent of the geocoding errors and only four percent of the correct enumerations. CIFU took an action on a relatively small percentage of the addresses because it occurred late in the MAF building process and also because the CIFU listers only visited addresses that were previously classified as vacant or nonexistent. Similar results are presented in Table B10 for the Outside the blue-line (Address Listing) areas.

Table B10. Comparison of CIFU and the A.C.E. Classification of Addresses as "Not a Housing Unit," "Geocoding Error," or "Correct Enumeration" Outside the Blue-Line (Address Listing)

A.C.E. Housing Unit Status	percent of addresses that CIFU took action on	percent inconsistent with CIFU	percent consistent with CIFU	unresolved percent consistent / inconsistent
"Not a Housing Unit" total addresses = 418,367 (weighted)	36.6 (33.6,39.6)	86.7 (83.5,89.9)	13.3 (10.2,16.5)	
"Geocoding Error" total addresses = 43,761 (weighted)	26.2 (15.3,37.1)	93.0 (86.4,100.0)	**	
Correct Enumeration total addresses = 20,122,938 (weighted)	9.6 (9.0,10.2)	NA	95.9 (95.4,96.4)	4.1 (3.6,4.6)

blank cells are cells that contain no addresses

B.6   Conclusions

Looking at the results, the percentage consistent for correct enumerations and percentage inconsistent for erroneous enumerations are high for most of the operations. It is reassuring to see the percentage consistent for correct enumerations as being high. The high percentage inconsistent for erroneous enumerations is also expected because these operations generally agree with the final census status over the final housing unit status in the A.C.E.

The percentage consistent of correct enumerations for each of the operations inside the blue-line is as follows:

- Block Canvassing   98.9 percent
- NRFU   97.6 percent
- CIFU   95.9 percent
- LUCA 98   88.9 percent
- LUCA 98 Field Verification 85.3 percent

Block Canvassing shows a significantly higher percentage consistent for correct enumerations than any of the other operations. NRFU is the next highest and is

significantly higher than CIFU, while CIFU is significantly higher than both LUCA 98 and LUCA 98 Field Verification. These results may suggest that an operation whose primary purpose is to update the address list shows better performance. The primary purpose of Block Canvassing was to update the address list. Updating the address list was a secondary purpose of NRFU and CIFU. It may stand to reason that the quality of updating the address list suffers a little in these operations because the field representatives had other, more important responsibilities. As stated in the results section, we believe LUCA 98 Field Verification showed relatively low percentage consistent because it did not have the ability to add addresses. The initial phase of LUCA 98 may have shown low percentage consistent because the updates generally were made from offices instead of from census field operations.

In the Outside the blue-line (Address Listing) areas, the three operations that show the highest percentage consistent are not significantly different from each other
• NRFU   96.6 percent
• Questionnaire Delivery   96.3 percent
• CIFU   95.9 percent
As explained in the results section, the percentage consistent for Address Listing and LUCA 99 Relisting may be low because these operations did not account for suffixed block codes. See section B.5.1 for a more detailed explanation about the suffixed block codes.

There are some interesting results dealing with the percent consistent for "not a housing unit" addresses. These addresses are the ones that A.C.E. and the MAF building operation agreed did not exist in the census block. It is interesting to see that the census, as a whole, showed many of the units as existing in the census blocks, while some of the individual census operations were trying to tell us that they were not existing in the blocks. Of the MAF building operations that we are examining, the highest percentages consistent for "not a housing unit" addresses are shown by
• Questionnaire Delivery   31 percent
• LUCA 98 Field Verification   26 percent
• CIFU   20 percent

A possible explanation why Questionnaire Delivery shows the highest percentage is that updates from the Update/Leave operation did not get incorporated into the MAF in time for the NRFU operation. NRFU did not tell us the correct block that a unit was in, so when the NRFU results were processed on the MAF, if a NRFU lister enumerated a unit it was assumed that the unit was enumerated in the correct

block. Because NRFU occurred after Update/Leave, a lot of NRFU actions overwrote Update/Leave actions in determining the final census status. We can see that when Update/Leave results were later incorporated into the MAF, it resulted in a lot of the updates that may have been able to correct block codes for the census.

A possible explanation for why LUCA 98 Field Verification has one of the highest percentage consistent for "not a housing unit" addresses among the operations is that this operation could not add addresses. If it would have had the ability to add addresses it may have added some to the census blocks, where a lot of other census operations added the addresses to. This would cause more LUCA 98 Field Verification actions to be inconsistent with the A.C.E. action and, therefore, cause more LUCA 98 Field Verification actions to be consistent with the A.C.E.

The same type of situation exists for the geocoding error addresses, in which Questionnaire Delivery and LUCA 98 Field Verification show the highest numbers for percentage consistent.

When comparing operations to each other, one should keep the following points in mind:

•   The different operations occurred at different time periods, which may affect
        the extent that an operation agrees with the final census status and with the
        A.C.E.

•   The different operations were of different sizes, and therefore had different
        impacts on the MAF, and

•   within an operation, one should be aware of the level of agreement between
        the operation and the A.C.E. for erroneous enumerations as well as correct
        enumerations

•   A.C.E. results can be incorrect. Therefore, we cannot use A.C.E. results to
        determine when a MAF building operation provides the right or wrong status
        for a unit. We can only compare the status from the MAF building operation
        to the status from the A.C.E.

•   An address can be acted upon by several different operations. Therefore,
        one address can be accounted for in several different operational tables in
        the Results section.

71

B7. References

Bureau of the Census, *Census 2000 Dress Rehearsal Evaluation Summary*, Washington, D.C., 1999.

"Census 2000 Type of Enumeration Areas (TEAs)," *Geography Intranet at the Bureau of the Census*, <http://www.geo.census.gov/mob/homep/teas.html> (March 26, 2001).

Childers, Danny R. *Accuracy and Coverage Evaluation: The Design Document*, Decennial Statistical Studies Division Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-1, January 24, 2001.

Furno, Gemma., *Census 2000 Telephone Questionnaire Assistance Program Master Plan:* Census 2000 Informational Memorandum No. 111, Bureau of the Census, August 14, 2001.

Gloster, John W., *Program Master Plan: Census 2000 Nonresponse Followup*, Census 2000 Informational Memorandum No. 26, Bureau of the Census, October 13, 1999.

Halterman, Kathleen., *Program Master Plan: Census 2000 Urban Update/Leave*, Census 2000 Informational Memorandum No. 88, Bureau of the Census, January 2, 2001.

Pike, A. Edward, *Program Master Plan: Census 2000 Master Address File*, Census 2000 Informational Memorandum No. 102, Bureau of the Census, May 1, 2001.

Pike, A. Edward, *Program Master Plan: Census 2000 Matching/Geocoding Non-Master Address File (MAF) Identification (ID) Questionnaires*, Census 2000 Informational Memorandum No. 98, Bureau of the Census, April 12, 2001.

Sanders, Monique V., *Program Master Plan: Coverage Improvement Followup*, Census 2000 Informational Memorandum No. 81, Bureau of the Census, November 17, 2000.

Vitrano, Frank A., *Determining Original Source for the November 2000 Master Address File for Evaluation Purposes*, Planning, Research, and Evaluation Division TXE/2010 Memorandum Series: MAF-EXT-S-01, Bureau of the Census, March 5, 2001.

Vitrano, Frank, et al., *Quality of Census 2000 Processes (Prototype)*, Decennial Statistical Studies Division Census 2000 Procedures and Operations Memorandum Series B-3, Bureau of the Census, September 27, 2000.

Walker, Shelley., *Program Master Plan: Census 2000 List/Enumerate Operation*, Census 2000 Informational Memorandum No. 46, Bureau of the Census, March 2, 2000.

Walker, Shelley, *Program Master Plan: Census 2000 Questionnaire Assistance Centers/Be Counted Program*, Census 2000 Informational Memorandum No. 103, Bureau of the Census, May 1, 2001.

Walker, Shelley., *Program Master Plan: Census 2000 Remote Alaska Program Master Plan*, Census 2000 Informational Memorandum No. 48, Bureau of the Census, March 8, 2000.

Walker, Shelley, *Program Master Plan: Census 2000 Update/Enumerate Program Master Plan*, Census 2000 Informational Memorandum No. 79, Bureau of the Census, November 9, 2000.

APPENDIX C. Technical Documentation

C.1   Housing unit dual system estimation output files

For the detailed file specifications and record layouts of the housing unit dual system estimation output files, see:

- Burcham, Joseph (2001) *"Accuracy and Coverage Evaluation: Creation of Input Files for Computing the Housing Unit Dual System Estimates"* Update to Q-55, (update 1 dtd 6/26/01)

- Hefter, Steven P. (2001) *"Accuracy and Coverage Evaluation: Housing Unit Dual System Estimation Programming Specifications(U.S.)*, Reissue of Q-56, (revised draft dtd 8/14/01)*

C.2 Variable Recodes and New Variables

B.2.1 For P-sample Processing (PHUFO_US)

OCCUP (Occupancy Status)

If FINOUTC in (1, 2, 3) then OCCUP = 1
Else if FINOUTC in (10, 11) then OCCUP = 2
Else OCCUP = 3

B.2.2 For E-Sample Processing (EHUFO_US)

OCCUP (Occupancy Status)

If FINST = 1 then OCCUP = 1
Else if FINST = 2 then OCCUP = 2
Else if FINST = 3 then do
    If NP = 0 then OCCUP = 2
    Else if NP ne 0 then OCCUP = 1

TENURE

If TENURE in (1, 2) then TENURE = 1
If TENURE in (3, 4) then TENURE = 2

ENUMTYPE

If DUP2 > 0 and FHICODE NOT = UE then ENUMTYPE = DE
Else if FHICODE = GU then ENUMTYPE = GE
Else if FHICODE in (P,MU,UE) then ENUMTYPE = UE
Else ENUMTYPE = FHICODE

C.3 Calculation of Percent P-sample Nonmatch

Filename: PHUFO_US.DAT
Variables: (PRHU, PRM, TRIMWTP, TESWGT)

Formula :   Nonmatch rate $=$ $1 - \dfrac{M}{Np}$ or $\dfrac{NMp}{Np}$

where:

M =   the weighted number of P-sample matched housing units, or

$$M = \sum_{k} PRM * TRIMWTP * TESWGT$$

NM_p =the weighted number of P-sample nonmatched housing units, or

$$NM_p = \sum_{k} (1 - PRM) * TRIMWTP * TESWGT$$

N_p =   the weighted number of P-sample housing units, or

$$N_p = \sum_{k} TRIMWTP * TESWGT$$

k =   the subset of housing units of interest; i.e., vacant, single unit, etc.

Percent P-sample nonmatch = nonmatch rate * 100

C.4 Calculation of Percent Erroneous Enumeration

Filename: EHUFO_US
Variables: (PRCE, DUPFACT, TRIMWTE, TESWGT, TES2WGT)

Formula: Erroneous Enumeration Rate $= 1 - \dfrac{CE}{N_e}$ or $\dfrac{EE}{N_e}$

where:

CE = the weighted estimate of the number of correct enumerations in the E-sample, or

$$CE = \sum_k PRCE2 * TRIMWTE * TESWGT * TES2WGT$$

Where PRCE2 = PRCE*DUPFACT

EE = the weighted estimate of the number of erroneous enumerations in the E-sample, or

$$EE = \sum_k (1 - PRCE2) * TRIMWTE * TESWGT * TES2WGT$$

N_e = the weighted number of E-sample housing units, or

$$N_e = \sum_k TRIMWTE * TESWGT * TES2WGT$$

Percent of Erroneous Enumerations = Erroneous Enumeration rate * 100

*Note when calculating Erroneous Enumeration rates one may save processing time by only processing records where ESAMP = 1.

C.5 Dual System Estimates

Filename1:   CHUFO_US.DAT
Variables: (TRIMWTE, FINST) for variables UBSA2, TENURE, MSATEA, DOMAIN, and REGION

Filename2:   POST_US
Variables: (CCWO) for the variable OCCUP

Formula: The formula for the dual system estimate of the population of HUs is:

$$ \text{DSE} \; = \; \frac{(C) \; (\frac{CE}{N_e})}{\frac{M}{N_p}} $$

where CE, N_e, M, N_p are defined as above and:

C =   the count of housing units in the census (*does not include reinstated units*) or for variables UBSA2, TENURE, MSATEA, DOMAIN and REGION, from CHUFO_US

$$ C = \; \sum_{k} \text{IND}(\text{FINST} \neq 3) $$

Where IND(statement) = 1 if the statement is true, 0 otherwise.

or, for variable OCCUP, from POST_US

$$ C = \; \sum_{k} \text{CCWO} $$

C.6 Determining reinstated units

Filename1:   CHUFO_US.DAT
Variables: (TRIMWTE, FINST) for variables UBSA2, TENURE, MSATEA, DOMAIN, and REGION

Filename2:   POST_US
Variables: (CCWO) for the variable OCCUP

Formula: For UBSA2, TENURE, MSATEA, DOMAIN and REGION, from CHUFO_US

$$\text{Reinstated Units} = \frac{\sum_k \text{IND}(\text{FINST} = 3)}{C}$$

Where C is defined above.

For OCCUP, from POST_US (CCWO)

$$\text{Proportion of reinstated units} = \frac{\sum_k \text{CCW} - \text{CCWO}}{\sum_k \text{CCWO}}$$

Percent of reinstated units = Proportion of reinstated units * 100

C.7 Calculation of Percent Net Undercount

Filename1: CHUFO_US.DAT
Variables: (TRIMWTE, FINST) for variables UBSA2, TENURE, MSATEA, DOMAIN, and REGION

Filename2: POST_US
Variables: (CCWO) for the variable OCCUP

Formula: Undercount rate = DSE minus the census count including reinstated units, divided by the DSE, or

$$\text{Undercount rate} = \frac{\text{DSE} - C^*}{\text{DSE}}$$

where DSE is defined in B.5 above and:

C^* = the count of housing units in the census (*includes reinstated units*) or for variables UBSA2, TENURE, MSATEA, DOMAIN and REGION, from CHUF_US

$$C^* = \sum_k 1$$

or for variable OCCUP, from POST_US

$$C^* = \sum_k \text{CCW}$$

Percent Undercount = Undercount rate * 100

79

C.8 Type of Erroneous Enumeration

Filename: EHUFO_US.DAT
Variables: DUP2, FHICODE

Formulas: Percentage of Erroneous Enumerations with ENUMTYPE of 'EE' =

$$\frac{ee\_sum}{EE}$$

Percentage of Erroneous Enumerations with ENUMTYPE of 'GE' =

$$\frac{ge\_sum}{EE}$$

Percentage of Erroneous Enumerations with ENUMTYPE of 'de' =

$$\frac{de\_sum}{EE}$$

Percentage of Erroneous Enumerations with ENUMTYPE of 'ue' =

$$\frac{ue\_sum}{EE}$$

where:

ee_sum = nationwide weighted estimate of the records with ENUMTYPE = 'EE', or

$$ee\_sum = \sum_{ee} (1 - PRCE2) * TRIMWTE * TESWGT * TES2WGT$$

ge_sum = nationwide weighted estimate of the records with ENUMTYPE = 'GE', or

80

$$\text{ge\_sum} = \sum_{ge} (1 - PRCE2) * TRIMWTE * TESWGT * TES2WGT$$

de_sum =   nationwide weighted estimate of the records with
ENUMTYPE = 'DE', or

$$\text{de\_sum} = \sum_{de} (1 - PRCE2) * TRIMWTE * TESWGT * TES2WGT$$

ue_sum =   nationwide weighted estimate of the records with
ENUMTYPE = 'UE', or

$$\text{ue\_sum} = \sum_{ue} (1 - PRCE2) * TRIMWTE * TESWGT * TES2WGT$$

EE =   nationwide weighted estimate of erroneous enumerations, or

$$EE = \sum (1 - PRCE2) * TRIMWTE * TESWGT * TES2WGT$$

PRCE2 = PRCE*DUPFACT

*Note that when calculating rates in this section, one may save processing time by only processing records where ESI = 1.

C.9 Identifying evidence of misclassification of vacant housing units

Create the dataset HUPER

Merge the E-sample HUDSE Input file, EHUFO, to the estimation file for census people EFINUS, by cluster and census ID (CLUST, CID)
In HUPER keep only the records where ESI = 1 and FHICODE = M
Drop records from EHUFO if there is no matching EFINUS record
From EFINUS keep only the variables CLUST, CID, and CEPROBF

Create the dataset MAX

Sort HUPER by CLUST, CID
Maintain the variable MAXCE from data step to data step (retain statement)
For the first person record in each CID by group set MAXCE = CEPROBF
For each subsequent person record in the CID by group if CEPROBF > MAXCE set MAXCE = CEPROBF
For the last person record in each CID by group output to MAX

Create the dataset MAXPER

Merge MAX with the reformatted CUF (only those records in the E-sample).
From the reformatted CUF keep the variables CLUST, CID, ESAMP, INPS, INP, FINST and SFINST
Output to MAXPER a record for every E-sample record that is on the reformatted CUF
If a record is on the reformatted CUF and in the E-sample, but not on MAX, assign an arbitrary value to MAXPER not between 0 and 1 (for example, 5)

Create the dataset VAC

Merge MAXPER with PHUFO_US by CLUST and FHICID from PHUFO_US to CLUST CID on MAXPER
Do not include in the merge any records on PHUFO_US with FINOUTC = 12
Output to VAC only those records for there is a CLUST, CID match from both files
Create variable ACEOCCUP
     If FINOUTC in (4, 6, 9) then ACEOCCUP = nonint
     If FINOUTC in (1, 2, 9) then ACEOCCUP = occupy
     If FINOUTC in (10, 11) then ACEOCCUP = vacant
Create variable CENVAC

If MAXCE=0 then do;
      If INP = 0 then CENVAC = vacant
      If INP > 0 then CENVAC = EE and II
If MAXCE = 5 and INP = 0 then CENVAC = finst?
If MAXCE = 5 and INP > 0 then CENVAC = II only
If MAXCE > 0 and MAXCE < 1 then CENVAC = unresolved
If MAXCE = 1 then CENVAC = CE people

Generate Table A.C.E. versus Census Occupancy Status for Matched E-Sample Units

From the dataset VAC produce crosstabulations of FINST with ACEOCCUP. Weight by the P-sample housing unit weights from the PHUFO_US, WEIGHTP and TESWGT

Generate Table Source of Final Status for E-Sample Units with only Non-Data Defined People

From the dataset VAC produce a tabulation of SFINST for the units with CENVAC = II only. Weight by the P-sample housing unit weights from PHUFO_US, WEIGHTP and TESWGT

83