

CENSUS 2000 HOUSING UNIT COVERAGE STUDY
Diane F. Barrett, Micheal Beaghen, Damon Smith and Joseph Burcham
Bureau of the Census, Washington DC 20233

Key words: Dual system estimates, Erroneous enumerations, Census misses

The Housing Unit Coverage Study

The Housing Unit Coverage Study (HUCS) measured the Census 2000 housing unit coverage using data from the Accuracy and Coverage Evaluation (A.C.E.) (Barrett et al., 2001). Dual system estimation was used to estimate a net undercount of housing units in the Census 2000. The study also examined 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 of the final address list containing all housing units existing in the United States on April 1, 2000.

In 1990, the Census Bureau used information obtained from its Post Enumeration Survey or PES to measure the housing unit coverage of the census (Childers, 1992 and 1993).

The 2000 Accuracy and Coverage Evaluation

The A.C.E. operations were similar to those in the 1990 PES. Both were an enumeration survey of housing units conducted independently of the census in a nationwide sample of block clusters. For Census 2000, the matching results between the A.C.E. and the census were used to determine the number of people and housing units missed and erroneously included in the census (Childers, 2000). This paper focuses on the coverage of housing units only.

Samples

The A.C.E. 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 A.C.E. block clusters on census day. The E-sample or enumeration sample was the housing units enumerated in the census in the same sample A.C.E. block clusters.

The P-sample was matched to the E-sample. Based on this match, nonmatched P-sample addresses were assigned as P-sample nonmatches, and matching addresses were assigned as correct enumerations or erroneous enumerations.

Correct vs Erroneous Enumerations

Correct enumerations were those P-sample housing

units that matched E-sample housing units within the block cluster. This means that the census address existed as a housing unit on census day and was correctly enumerated in the block cluster.

There were several reasons for erroneous enumerations. The P-sample address may have matched to a census address but the address did not exist as a housing unit on census day or the address did exist as a housing unit on census day but was found outside the block cluster, that is, the housing unit was incorrectly assigned to the A.C.E. block cluster. Another type of erroneous enumeration was a duplicate unit of a housing unit already enumerated in the census.

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 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*)

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

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 housing unit inventory. None of these units were in the A.C.E. universe, thus were not used in the dual system estimation. However, these reinstated housing units

This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau Publications. This report is released to inform interested parties of research and to encourage discussion.

were included in the synthetic estimation for calculating the net undercount of housing units. Refer to Hefter, 2001 and Burcham, 2001.

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 one minus the ratio of the DSE minus the census count (including reinstated units), divided by the DSE. This ratio is then multiplied by 100 for the percent net undercount.

$$Pct\ Net\ Undercount = \frac{DSE - C^*}{DSE} \times 100$$

where: DSE = the formula defined above, C* = the count of housing units in the census (includes reinstated units).

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

Percent P-Sample Nonmatches. Census omissions are determined by the percent of housing units in the P-sample not matched to the census. 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 to give the percent of P-sample nonmatches.

$$Pct\ P\text{-sample}\ nonmatches = 1 - \frac{M}{N_p} \text{ or } \frac{NM_p}{N_p} \times 100$$

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

$$Pct\ erroneous\ enumerations = 1 - \frac{CE}{N_e} \text{ or } \frac{EE}{N_e} \times 100$$

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.

Post-Stratification and Research Categories

This report examines the housing unit coverage estimates by various research categories for the Census 2000 and where available, compares the 2000 estimates to the 1990 coverage estimates.

Five post stratification variables were used in the dual system estimation: occupancy status, census region, race/Hispanic origin, size of structure, and Metropolitan Statistical Area/Type of Enumeration (MSA/TEA) group. We also investigated coverage estimates at the national level and by tenure.

Single cell DSEs versus Production DSE's. For those research categories that were combined in one or more post-strata or where the research category was not a post-stratum variable (tenure), we calculated the net coverage estimate using a single cell DSE. The single cell DSE was used to calculate the net coverage for all research categories except national and occupancy status. For these research categories, we obtained the production DSE by summing over the appropriate post-strata.

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.

RESULTS

What was the national net coverage of housing units? See Table 1 for the national net percent undercount by occupancy status for 2000 and 1990.

Both censuses had a net percent undercount of less than 1.0 percent. The net undercount of housing units in the Census 2000 was 0.61 percent which was not significantly different from the net undercount in 1990 (0.96 percent). For occupied housing units, no significant difference was observed between the 2000 and 1990 estimates. The net undercount was 0.33 percent in 2000 and 0.53 percent in 1990. The net undercount for vacant units was 3.37 percent in 2000 which was not significantly different from the 4.71 percent net undercount in 1990. In 2000, vacant units (3.37 percent) were significantly undercounted more than occupied units (0.33 percent) which results in a difference of 3.04 percentage points. In 1990, the difference between the net undercount for vacant units (4.71percent) and occupied units (0.53 percent) was 4.18 percentage points. This difference was also significant.

We estimated the net undercount of housing units for Census 2000 as 0.61 percent, however, the true net undercount percent for the nation was between 0.35 and 0.87 percent. For occupied units, the true net undercount is somewhere between 0.12 and 0.54

percent. For vacant units the interval is a lot wider, somewhere between two and five percent.

What percentage of housing units did the census miss? The Census 2000 missed housing units at the same rate as in 1990. Table 2 shows the rate of omission was less than 4 percent. For both censuses, vacant units were missed more often than occupied units. In 2000, the difference between the estimates was a significant difference of about 11 percentage points. In 1990, the difference was about 10 percentage points. 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 the census in error.

What percentage did the census enumerate in error? See Table 3. About two percent of the housing units enumerated in the Census 2000 were enumerated in error. In 1990, it was about 3 percent. No significant difference was observed between the 2000 and 1990 estimates. The census erroneously enumerated vacant housing units more than occupied housing units in both the 2000 and 1990 censuses. For both censuses, the differences between occupied and vacant units were significant. Here again, applying the housing unit definition without the benefit of a respondent was difficult for census enumerators, thus these types of units may have been included in the census in error.

What was the major reason for erroneous enumerations of housing units? The major reason for classifying a unit as an erroneous enumeration 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 addresses have been combined into one type of erroneous enumeration category as “not a housing unit” (See Table 4).

Over half (57.05 percent) of all erroneous enumerations were not housing units. In 1990, not a housing unit (37.3 percent) and duplicates (33.4 percent) were both major reasons for erroneous enumerations. Duplicates in both the 1990 and 2000 censuses accounted for a large portion of the erroneous enumerations.

What were the coverage estimates by census region? Table 5 provides the Census 2000 coverage estimates for each census region.

The undercount in the Midwest improved by about one percentage point in Census 2000. The undercount in the Midwest was significantly lower in 2000 (0.19 percent) than in 1990 (1.13 percent). Undercount in the other regions in Census 2000 was not significantly different than in 1990. In 2000, more housing units were missed in the Northeast (4.23 percent) than in the Midwest (2.67 percent). More housing units were missed in the South than in the Midwest in both 2000 and 1990. In 2000, housing units in the South (2.58 percent) were enumerated in error more often than housing units in the Midwest (1.8 percent). In both censuses, housing units in the Northeast (about 3 percent in each census) were enumerated in error more often than housing units in the Midwest (1.2 percent in 1990, 1.8 percent in 2000).

What were the coverage estimates by tenure? Table 6 compares coverage estimates of occupied housing units between owner and renter. Whether a housing unit was owner occupied or renter occupied had no impact on housing unit coverage. No significant differences in the undercount were observed between housing units with owners and renters.

What were the housing unit coverage estimates by race/Hispanic origin of householder for occupied housing units? We analyzed the coverage estimates of housing units by race/Hispanic origin of the householder as shown in Table 7. 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 same domain of the householder, that is person 1 on the Census questionnaire.

The percent net undercount (-0.45 percent) for housing units with Non-Hispanic Black householders was significantly lower than the estimated net undercount of 0.38 percent for housing units with Non-Hispanic White or Some Other Race householders. However, their nonmatched and erroneously enumerated percentages were not different.

Coverage of Hispanic housing units and Non-Hispanic White or Some other race was not significantly different.

What were the housing unit coverage estimates by size of structure? Table 8 examines coverage estimates by the three sizes of structure categories defined for 2000. The size of structure is based on the number of units at the basic street address. The three sizes for Census 2000 were single units with one unit at the basic street address, small multiunit structures with 2 to 9 units at the basic street address and large multi-unit structures with 10 or more units at the basic street address.

In 1990 there were five types of structure categories. We also used the number of units at the basic street

address as a proxy for type of structure in 1990. The categories were single units, small multiunits with 2 to 9 units, medium multiunit with 10 to 49 units, large multiunits with 50 or more housing units, and an “other” category which included mobile homes. In Table 8 we only compared the net percent undercount between 2000 and 1990 for single units and small multi-unit structures, since the large multi-unit category was different. Also, note that the 2000 categories include mobile homes whereas in 1990, mobile homes were included in a separate “other “ category.

No significant differences of net percent undercount was observed by size of structure in 2000. Small multiunits had a net undercount of -0.17 percent which was not significantly different from zero. Small multiunits were overcounted in 2000 but were significantly undercounted in 1990 at 2.25 percent. The overcount for small multiunits (-0.17 percent net undercount) was also significantly different than the coverage for single units (0.76 percent net undercount) but not significantly different from large multiunits. Large multiunits had a net undercount of -0.13 percent which was not significantly different from zero. Coverage of small multiunit structures was the most problematic among the three sizes of structures for Census 2000. The percent of P-sample nonmatches (6.94 percent) and the percent of erroneous enumerations (4.78 percent) for small multiunits were both significantly higher than for single units and for large multiunits.

What were the coverage estimates by Metropolitan Statistical Area/Type of Enumeration Area (MSA/TEA) group? We have calculated coverage estimates by size of metropolitan area and type of enumeration method used for Census 2000. See Table 9.

Size of metropolitan area had no impact on the percent net undercount in mailout/mailback areas. The estimates were about the same. However, when we examined percent of P-sample nonmatches we found that size made a difference. We missed more housing units in small metropolitan areas (category also includes non-metropolitan areas) than medium metropolitan areas in mailout/mailback areas. For erroneous enumerations, medium MSAs had the lowest percentage when compared to the other mailout/mailback areas.

Conclusion

The overall coverage of housing units in the Census 2000 was similar to 1990. Both censuses resulted in a net undercount of less than one percent; they missed less than 4 percent of the housing units; and erroneously enumerated about 2 to 3 percent of the housing units.

Coverage of housing units improved in some research categories. The undercount in the Midwest improved

by one percentage point. Coverage of small structures with 2 to 9 units was significantly lower in 2000 (-0.17%) than in 1990 (2.25%). In the Census 2000, occupied housing units with Non-Hispanic Black householders had better coverage than housing units with Non-Hispanic White and Some other race householders.

Although net coverage of housing units in small multiunits improved significantly over 1990, small multiunits had the highest percent of P-sample nonmatches and erroneous enumerations compared to the other type of structures. Correctly enumerating vacant units continues to be an issue for the census. Estimates of net coverage, misses and erroneous enumerations was significantly greater for vacant units than occupied units.

References

- Barrett, Diane, Beaghen, Micheal, Smith, Damon and Burcham, Joseph** (2001) “*Executive Steering Committee on Accuracy and Coverage Evaluation Policy II Report Number 17: Census 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 Study*” Presented at the American Statistical Association Meetings in August 1992 in Boston, Massachusetts. 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.** (2001) “*Accuracy and Coverage Evaluation: Housing Unit Dual System Estimation Programming Specifications (U.S.)*”, DSSD Census 2000 Procedures and Operations Memorandum Series Reissue Q-56, Bureau of the Census, August 14, 2001.

Table 1. National Percent Net Undercount of Housing Units By Occupancy Status* (s.e)

Status	2000 ACE	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)

*used production DSEs

Table 2 . National Percent of P-sample Nonmatches by Occupancy Status (s.e.)

Status	2000 ACE	1990 HUCS
National	3.62 (0.15)	3.57 (0.2)
Occupied	2.61 (0.11)	2.54 (0.2)
Vacant	13.54 (0.79)	12.67 (1.0)

Table 3. National Percent of Erroneous Enumerations (s.e)

	2000 ACE	1990 HUCS
National	2.31 (0.11)	2.84 (0.20)
Occupied	1.51 (0.07)	2.17 (0.10)
Vacant	10.50 (0.67)	9.24 (1.0)

Table 4. Percent of Erroneous Enumeration by Reason (s.e.)

Reason	2000 Percent	1990 Percent
Duplicates	24.81 (2.76)	33.4 (na)
Geocoding errors	16.15 (1.72)	16.2 (3.0)
Not a housing unit	57.05 (2.51)	37.3 (3.4)
Unresolved	1.99 (0.56)	2.8 (0.4)
Insufficient Information	na	10.2 (2.0)

na-not applicable

Table 5. Housing Unit Coverage Rates by Census Region (s. e.)

Census Region	Percent P-Sample Nonmatches		Percent Erroneous Enumeration		Net Percent* Undercount	
	2000 A.C.E.	1990 HUCS	2000 A.C.E.	1990 HUCS	2000* A.C.E.	1990 HUCS
Northeast	4.23 (0.34)	4.1 (0.5)	2.73 (0.21)	3.0 (0.5)	.47 (0.40)	0.53 (0.52)
Midwest	2.67 (0.22)	2.6 (0.4)	1.80 (0.14)	1.2 (0.2)	0.19 (0.26)	1.13 (0.43)
South	3.92 (0.26)	3.9 (0.2)	2.58 (0.23)	2.2 (0.4)	0.44 (0.28)	0.80 (0.43)
West	3.58 (0.35)	2.8 (0.5)	2.04 (0.19)	1.6 (0.3)	0.86 (0.40)	1.48 (0.58)

*used single cell DSEs

Table 6. Housing Unit Coverage Estimates by Tenure (s.e.)

Tenure	Percent P-Sample Nonmatches	Percent Erroneous Enumerations	Percent * Net Undercount
Owner-occupied	2.14 (0.11)	1.26 (0.07)	0.12 (0.13)
Renter-occupied	3.56 (0.22)	2.02 (0.15)	0.57 (0.26)

*used single cell DSEs

Table 7. Housing Unit Coverage Estimates by Race/Hispanic Origin of Householder in Occupied Units (s.e.)

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

*used single cell DSE na-not available

Table 8. Housing Unit Coverage Estimates by Type of Structure (s.e)

Size of Structure	Percent P-Sample Nonmatches	Percent Erroneous Enumeration	Percent Net Undercount*	
			2000* A.C.E.	1990 HUCS
Single Units	3.18 (0.15)	1.78 (0.07)	0.76 (0.16)	0.76 (0.23)
Small Multiunits 2 to 9 HUs	6.94 (0.57)	4.78 (0.23)	-0.17 (0.64)	2.25 (0.65)
Large Multiunits 10 or more HUs	3.39 (0.44)	2.97 (0.51)	-0.13 (0.54)	na

*used single cell DSE

Table 9 . Housing Unit Coverage Estimates by Metropolitan Statistical Area/Type of Enumeration Area (MSA/TEA) Group (s.e.)

MSA/TEA	Percent P-Sample Nonmatches	Percent Erroneous Enumeration	Percent Net* Undercount
Large MSA Mailout/Mailback	3.01 (0.24)	2.13 (0.17)	0.22 (0.29)
Medium MSA Mailout/Mailback	2.41 (0.22)	1.60 (0.14)	0.41 (0.25)
Small MSA&NonMSA Mailout/Mailback	3.59 (0.34)	2.62 (0.38)	0.58 (0.35)
All Other TEAs	6.52 (0.44)	3.38 (0.14)	1.01 (0.47)

*used single cell DSE