Enumeration Status of E-Sample Duplicates Outside the Accuracy and Coverage Evaluation Search Area for Census 2000

Roxanne Feldpausch and Vincent Mule[1]
Bureau of the Census, Washington, DC 20233

Keywords: census duplication, census residence rules, erroneous enumerations

1.0 Background

In past censuses, there was no way to evaluate the post-enumeration survey's coding of people duplicated outside its search area. For Census 2000 there was an analysis project that searched for duplicates throughout the country. The Census Person Duplication Operation matched person records across the country to determine the extent of duplication. We used the results of the Census Person Duplication Operation to evaluate the Accuracy and Coverage Evaluation (A.C.E.) coding of people. This paper looks at the A.C.E. coding of people duplicated outside the A.C.E. search area.

Sections 1.1 and 1.2 give some background on the A.C.E. and Census Person Duplication Operation, respectively. Section 2 discusses the methods used for this paper. Section 3 list the limitations. Section 4 gives the results. Section 5 summarizes the findings and gives possible explanations for the findings.

1.1 The Accuracy and Coverage Evaluation

The A.C.E. was an operation undertaken to evaluate the coverage of Census 2000. It was comprised of the matching of an independent enumeration of housing units and people in a sample of census block clusters against the Census 2000 enumerations in those block clusters. The A.C.E. included an initial housing unit phase, a person interview phase, a person match phase, and a final housing unit phase. For more information on the A.C.E. see Childers (2001).

The A.C.E. person matching was conducted in the fall of 2000, after the A.C.E. person interview phase was complete. During A.C.E. person matching E-Sample people were matched to the people independently enumerated by the A.C.E. The E-Sample was a sample of census housing units in A.C.E. block clusters and their corresponding people. The A.C.E. person matching consisted of three stages: computer matching, before follow-up clerical matching and after follow-up clerical matching. During A.C.E. person matching, match and enumeration codes were assigned to census person records based on the following characteristics:

- Complete Name (First, Middle Initial, and Last)
- Date of Birth
- Age
- Sex
- Race/Hispanic origin
- Tenure

The A.C.E. clerical matchers were able to look at the other people in the household when making coding decisions. They conducted a search for duplicate census person records. The A.C.E. search for duplicates occurred within housing units in the block cluster; group quarters were not included in the A.C.E. search for duplicates. In addition, a sample of block clusters had their search area expanded to include the first ring of surrounding blocks.

Based on the results of A.C.E. person matching, an E-Sample person could be considered correctly or erroneously enumerated. The people that the census captured correctly within the A.C.E. search area were coded as correctly enumerated. There were two types of correct enumerations:

- **Matches**: The census and A.C.E. both captured the same person within the A.C.E. search area.
- **Correct enumerations**: The census captured the person, but the A.C.E. did not capture the person. Based on the results of a follow-up interview, it was determined that the census captured the person in the correct block cluster.

The people that the census captured in error in the block cluster were coded as erroneous enumerations. There are five types of erroneous enumerations; two of them are relevant to this paper.

- **Duplicate**: The census captured the person more than once within the A.C.E. search area.

[1] This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

- **Other residence**: The census counted the person in error, because the person should have been counted at another residence outside the A.C.E. search area. People duplicated outside the A.C.E. search area should have been coded as erroneously enumerated due to other residence.

People classified as correct or erroneous enumerations based on A.C.E. person matching and follow-up were assigned a correct enumeration probability. If we could not determine the person to be correctly or erroneously enumerated, the enumeration status for the E-sample person was unresolved. Those people with unresolved enumeration status had their probability of correct enumeration imputed.

1.2 The Census Person Duplication Operation

The research into person record duplication across the country was brought about in part to evaluate the Housing Unit Duplication Operation. The Housing Unit Duplication Operation was a two stage process to remove duplicate housing units and the people in them from the census. The first stage identified 5.9 million person records as potential duplicates and flagged them for possible deletion. These person records were temporally removed from the census; they were also excluded from the E-Sample universe. The second stage analyzed the potential duplicates to determine which ones were actual duplicates. During the second stage, 2.3 million person records were reinstated back into the census (they were still excluded from the E-Sample universe). The remaining person records were deleted from the census. For details on the Housing Unit Duplication Operation, see Nash (2000).

The Census Person Duplication Operation looked for duplicated people across the country. It looked for duplicates in source records (person records in E-Sample eligible housing units and reinstated units in A.C.E. block clusters). The source records were matched to target records (census person records and person records deleted from the census during the Housing Unit Duplication Operation). The Census Person Duplication Operation used a two stage computer matching program to identify duplicate records across the entire country. The first stage of matching used an exact matching procedure. For two person records to be a match during this stage, the following variables were required to match exactly:

- First Name
- Last Name
- Month of Birth
- Day of Birth

The second stage used statistical-based matching. The second stage matching was only done in housing units where there was an exact match identified in the first stage of matching. In addition to the variables used in the first stage, the second stage used middle initial and computed age. See Mule (2001) for details on the Census Person Duplication Operation.

Two factors were applied to each link identified during the Census Person Duplication Operation:

- An unbiased probability of duplication for the link
- A model weight which expressed the confidence in the link representing true duplication

2. Methods

Duplicate links where an E-Sample person record was the source and the target unit was outside the surrounding blocks were used to evaluate the A.C.E. person matching. These data came from the Census Person Duplication Operation. The unbiased probabilities of duplication were computed based on E-Sample person records being the only records eligible to be a source record. We only considered Census Person Duplication links that had a model weight greater than 0.5, in other words we only considered duplicate links that we had confidence in.

All records were weighted with the final E-Sample person weight from the A.C.E. and with the unbiased probability of duplication for an E-Sample person record. The records were not weighted with the model weight. Person records were classified as match, correct enumeration and erroneous enumeration based on their correct enumeration probabilities and A.C.E. match codes. Standard errors were produced in VPLX using simple Jackknife.

Note that the unbiased probability of duplication appropriate for this paper and the use of the model weight cause the numbers presented here to differ from those in Mule (2001).

3. Limitations

This report does not separately examine the issue of movers. It does not measure separately the amount of duplication due to people moving during the time frame the enumeration took place.

This report assumes that the duplicates identified by the Census Person Duplication Operation are correct. This report only considers cases where the Census Person Duplication Operation team has confidence in the link. The report does not examine the quality of the Census Person Duplication Operation.

4. Results

E-Sample people identified by the Census Person Duplication Operation as duplicates to people outside the surrounding blocks are the focus of this paper. Table 1 shows the number of E-Sample people duplicated to people outside the surrounding blocks by the type of unit of the target. The target unit can be an E-Sample eligible housing unit, a group quarters, a reinstated unit, or a deleted unit. The reinstates and deletes were those units that were flagged during the Housing Unit Duplication Operation.

Table 1 Number of E-Sample People Duplicated to People Outside the Surrounding Blocks

Target: Type of Unit	Total	(Standard Error)
Housing Unit	2,169,366	(49,926)
Group Quarters	520,501	(37,069)
Reinstates	574,169	(30,959)
Deletes	264,662	(22,117)

For E-Sample people duplicated to people in housing units we expected about half to be erroneous enumerations. Because the A.C.E. was a random sample, one would expect that about half the time, the A.C.E. would capture the person record in the correct housing unit according to census residence rules. These person records should have been coded correctly enumerated or matched by the A.C.E. One would expect the other half of the time, the A.C.E. would capture the person record in the housing unit that was incorrect according to census residence rules. These person records should have been coded as erroneously enumerated by the A.C.E.

Table 2 shows the percent match, percent correct enumeration and percent erroneous enumeration of E-Sample people duplicated to people outside the surrounding blocks. Table 2 breaks down the target people based on type of unit in which they lived. Group Quarters are excluded from this table. They will be discussed separately.

About 14.2 percent of the E-Sample people duplicated to people in housing units outside the surrounding blocks were erroneously enumerated. This is lower than the 50 percent we expected based on the fact that the A.C.E. was a random sample. The implications of this result are discussed in Section 5.

Table 2 shows that 59.8 percent of the E-Sample people duplicated to people in housing units outside the surrounding blocks were matched. The high match rate was also a cause for concern. A matched person is someone who was captured independently by the A.C.E. and the census. The A.C.E. was conducted via personal interview using a Computer Assisted Personal Interview (CAPI) instrument. It contained probing questions that were supposed to identify people who could have been counted someplace else. We expected that the A.C.E. would do a better job than the census at identifying people who should have been counted elsewhere.

We expected the coding of E-Sample people duplicated in reinstated and deleted units to be similar to that of people duplicated in housing units. If the Housing Unit Duplication Operation favored geographically correct units, we expect the percent erroneous enumeration to be lower for people duplicated in reinstated and deleted units.

Table 2 Percent Match, Correct Enumeration and Erroneous Enumeration of E-Sample Duplicates (Standard Error)

	HU	Reinstate	Delete
Match	59.8 (0.9)	48.0 (2.6)	46.1 (4.0)
Correct	26.0 (0.7)	33.4 (2.4)	34.5 (3.7)
Erroneous	14.2 (0.7)	18.6 (1.9)	19.4 (3.5)

Group Quarters can be split into two categories: those that could claim usual home elsewhere (UHE) and those that could not claim UHE. Census residence rules state that people counted in certain types of group quarters should be counted in the place where they live and sleep most of the time. People in these types of group quarters can claim UHE. Group quarters that can claim UHE include military barracks, worker dorms and people enumerated in soup kitchens. Census residence rules state that people staying in other types of group quarters should be counted in the group quarters and they cannot claim UHE.

Table 3 shows the number of E-Sample people duplicated to people in group quarters outside the surrounding blocks. Table 3 splits E-Sample people duplicated in group quarters into two categories: those that could claim UHE and those that could not claim UHE. Those that could not claim UHE were split into two categories: those in a college dorm and those not in a college dorm.

Table 3 Number of E-Sample People Duplicated to People in Group Quarters Outside the Surrounding Blocks

Target: Type of Unit	Total	(Standard Error)
Could not claim UHE - Not a Dorm	189,756	(11,028)
Could not claim UHE - Dorm	271,158	(34,806)
Could claim UHE	59,586	(5,915)

The census forms for group quarters where the respondent could claim UHE contained a box to check if the respondent wanted to claim UHE and a space for the address of the usual residence. During A.C.E. person matching, the clerical workers did not have access to information on group quarters forms from across the country. General coding rules were developed for cases where the A.C.E. person interview indicated the person spent some time in a group quarters, but was also counted in the housing unit. Childers (2001) shows how an E-Sample person should have been coded when the A.C.E. determined that the person should have also been counted in a group quarters:

If the person lived in a GQ where they...	the E-Sample person should have been coded as...
could claim UHE	correctly enumerated
could not claim UHE	erroneously enumerated

Table 4 shows the top five group quarters that account for about 80.5 percent of the E-Sample people duplicated to people in group quarters outside the surrounding blocks. College dorms accounted for 52.1 percent of these cases. Local jails account for 8.5 percent of the duplicates to group quarters. Some people counted in local jails and other temporary group quarters may actually have been usual residents of the sample housing unit, because they live and sleep there most of the time. However, according to census residence rules the person should have been counted in the local jail. Based on A.C.E. person matching procedures, the person in a local jail should have been coded as an erroneous enumeration if they were in the local jail on April 1, 2000. Table 4 also indicates whether or not the people in the group quarters could claim UHE.

Table 4 Percent of E-Sample people duplicated to Group Quarters by type of Group Quarters

	Percent	Is UHE OK?
College Dorm	52.1	No
Nursing Home	10.9	No
Local Jail	8.5	No
State Prison	4.5	No
Military Barracks	4.5	Yes
All Other GQ	19.5	

Table 5 shows the percent match, percent correct enumeration and percent erroneous enumeration of E-Sample people duplicated to people in group quarters outside the surrounding blocks. Assuming the group quarters enumeration was perfect, one would expect that almost all of those people duplicated to group quarters that could not claim UHE would be coded as erroneous enumerations. Table 5 shows that 45.4 percent of people duplicated to college dorms were coded as erroneous enumerations and 16.5 percent people duplicated to other group quarters that could not claim UHE were coded as erroneous enumerations. Assuming that the group quarters enumeration was perfect, one would expect that almost all of those people duplicated to group quarters that could claim UHE would be coded as matches or correct enumerations. Table 5 shows that 87.5 percent were coded as matches or correct enumerations.

Table 5 Percent Match, Correct Enumeration and Erroneous Enumeration of E-Sample Duplicates (Standard Error)

	Could not claim UHE		Could claim UHE
	Not a Dorm	Dorm	
Match	54.7 (2.8)	34.6 (7.4)	67.8 (4.2)
Correct	28.8 (2.6)	19.9 (2.0)	19.7 (3.8)
Erroneous	16.5 (1.9)	45.5 (6.6)	12.5 (2.6)

5. Conclusions

The percent of people in E-Sample eligible housing units who were erroneously enumerated was lower than expected. For these people, the percent match was higher than expected. For people in group quarters that could

not claim UHE, the percent erroneous enumeration was lower than expected. Based on these results, we conclude:

A.C.E. person matching did not correctly code many E-Sample records identified as duplicates in the Census Person Duplication Operation.

Possible explanations include:

- The instructions indicating who to include on the census questionnaire were not completely understood or the instructions may have been understood, but were ignored by the respondent. Examples include college students, people in local jail and people in nursing homes.

- The respondent may have assumed that the census residence rules were the same as residence rules for other government agencies, such as the IRS.

- The respondent may not have realize that a household member was enumerated elsewhere.

- Some group quarters' enumeration may have been done using administrative records that did not reflect the residents as of April 1, 2000.

- Some group quarters are temporary, such as local jails. Some people counted here may actually be usual residents of the sample housing unit.

- The computer matching of duplicates outside the search area might be incorrect. We do not believe that this was a large part of the explanation, because we only looked at those cases that we had confidence in (those cases that had a high probability of being linked correctly).

- The A.C.E. did not do an optimal job of identifying people who should have been coded as erroneous enumerations due to other residence. The percent other residence (after the redistribution of people with unresolved status) was 1.4 in the 2000 A.C.E. and 2.3 in the 1990 PES (Feldpausch 2001). The results of the Measurement Error Reinterview (Adams 2001) also measure this phenomenon.

6. References

Adams, T and Krejsa, E. A. (2001). "Results of the Person Followup and Evaluation Followup Forms Review," Executive Steering Committee on Accuracy and Coverage Evaluation Policy II Report Number 24, DSSD Census 2000 Procedures and Operations Memorandum Series T-17, Oct. 16, 2001, U.S. Census Bureau.

Childers, D. R. (2001), "Accuracy and Coverage Evaluation: The Design Document", DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-1, Revised, Jan. 26, 2001, U.S. Census Bureau.

Feldpausch, R. (2001), "E-Sample Erroneous Enumeration Analysis," Executive Steering Committee on Accuracy and Coverage Evaluation Policy II Report Number 5, DSSD Census 2000 Procedures and Operations Memorandum Series T-11, Oct. 10, 2001, U.S. Census Bureau.

Mule, T. (2001), "Person Duplication in Census 2000," Executive Steering Committee on Accuracy and Coverage Evaluation Policy II Report Number 20, DSSD Census 2000 Procedures and Operations Memorandum Series Q-71, Oct. 11, 2001, U.S. Census Bureau.

Nash F. (2000), "Overview of the Duplicate Housing Unit Operations," internal document, November 7, 2000, U.S. Census Bureau.