UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

March 28, 2012

DSSD 2012 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT MEMORANDUM SERIES #ACS-12-RER-15

DSSD 2012 AMERICAN COMMUNITY SURVEY MEMORANDUM SERIES #ACS12-UC-03

MEMORANDUM FOR ACS Research and Evaluation Steering Committee

From: David C. Whitford ***/Signed/***
Chief, Decennial Statistical Studies Division

Prepared by: Dianne Aubuchon
Community Address Update System
Decennial Statistical Studies Division

Subject: Evaluation of Decision to Use Ungeocoded Delivery Sequence File Records in Some Counties Completely Outside the Census 2000 Blue Line

Attached is the final American Community Survey Research and Evaluation report for the Evaluation of Decision to Use Ungeocoded Records in Some Counties Completely Outside the Census 2000 Blue Line. In the 2009 American Community Survey frame, we began to use ungeocoded Delivery Sequence File records in 63 counties where we previously did not use them. After these records had had a chance to be in sample for the American Community Survey, we evaluated the decision to begin using them.

If you have any questions about this report, please contact Jim Hartman (301-763-1976) or Dianne Aubuchon (301-763-7706).

Attachment

cc:
ACS Research and Evaluation Team
Tony Tersine (DSSD)
Jim Hartman
Mark Asiala
Steven Hefter
Scott Boggess (SEHSD)
Matt Zimolzak (ACSO)

March 28, 2012

Evaluation of Decision to Use Ungeocoded Delivery Sequence File Records in Some Counties Completely Outside the Census 2000 Blue Line

FINAL REPORT

United States™ Census Bureau

Dianne Aubuchon
Decennial Statistical Studies Division

1. Executive Summary

The American Community Survey has been using ungeocoded Delivery Sequence File units in areas that were either completely or partially mailout/mailback for Census 2000. In 2009, based on an indication of growth from the Population Estimates Program in areas where there was no mailout/mailback, the American Community Survey began to include ungeocoded Delivery Sequence File units in certain counties with a small percentage of non city-style addresses and from fast growing counties with more than 10 percent undercoverage.

A quick decision was needed on whether to continue allowing these records to be in the American Community Survey universe or not. We wanted to determine if the new records resulted in interviews.

- For the January – May 2009 American Community Survey panels, the added ungeocoded records resulted in an overall weighted total interview ratio of 79.3 percent, which, while lower than the panels as a whole, is still a good overall ratio.

- The mail check-in ratio and Computer Assisted Personal Interviewing interview ratio were similar between the added ungeocoded records and all ungeocoded records in the January - May 2009 panels.

- The likelihood of duplication in these records appears to be low. Of the records that resulted in an interview and had an Address Characteristic Type code, 95 percent were found to be in a block where at least 90 percent of units have a city style address.

- To assess if we should not use ungeocoded addresses in some areas, we focused on counties with an interview ratio of less than 50 percent. Given the fairly small sample sizes in these counties and the good overall interview ratio, we decided to continue using ungeocoded records in all of the areas studied.

2. Introduction

Between Decennial Censuses there are 2 main sources of updates to the Master Address File (MAF)/Topologically Integrated Geographic Encoding and Referencing database (TIGER) or MTdb. The first source is the U.S. Postal Service's (USPS) Delivery Sequence File (DSF), which is the primary source of city-style[1] addresses updates for the MTdb. The second source is the Demographic Area Address Listing (DAAL) operation, which is the primary source of non city-style address updates for the MTdb.

Since the 2000 Census, the American Community Survey (ACS) has used geocoded DSF units in areas that are within the 2000 blueline[2]. Geocoded units are those that are assigned to an identifiable location, such as a Census block. If a unit is not assigned to an identifiable location, then it is labeled as ungeocoded. ACS has been using ungeocoded DSF units in areas that were either completely or partially mailout/mailback for Census 2000.

[1] An address is marked city-style if it has a house number and street name; otherwise it is marked as non city-style.
[2] These are areas that were mailout/mailback in Census 2000.

The Population Estimates Program (PEP) indicates that there is a lot of growth in many counties where there were no mailout/mailback areas for the 2000 Census. These areas contained a significant number of non-city style addresses in 2000; if ungeocoded records are used, they would increase the risk of duplication. ACS is not picking up much of the growth in these counties.

A few years ago, ACS began using the Address Characteristic Type (ACT) code to expand our use of the DSF. The ACT is a code assigned to tabulation blocks. It gives us the make-up of addresses in that block, namely if they are city-style, non city-style, or mixed and whether or not there is DSF coverage in that block. The ACT code allows us to determine where there are areas that were non city-style that have now changed to city-style. This allowed us to start using geocoded new DSF records in areas that were outside the blueline, but contain many city-style addresses. It did not help with ungeocoded DSF addresses, as many new DSF addresses are ungeocoded.

Counts of valid ACS addresses from the July 2007 ACS MAF extracts were compared to the 2007 PEP estimate of housing units (HU) to determine where ACS might have a coverage problem. The ACS extracts contain select data from the MTdb used for ACS sample selection. The ACS filter[3] was rerun while ignoring the 2000 blueline status (which caused ungeocoded records in counties where they were previously excluded to be included), and the coverage rate recalculated. From this we saw that the coverage rate in certain areas improved by including these ungeocoded records.

ACS decided to start using certain ungeocoded DSF records in counties where either

1) The percentage of non city-style addresses is less than five percent (56 counties) or

2) The percentage of non city-style addresses is between five and 10 percent, the 2008 universe coverage rate is less than 90 percent, and the growth rate between Census 2000 and 2007 (per the PEP HU estimates) is more than 10 percent (seven counties)

Where the 2008 universe coverage rate was calculated as:

$$\frac{HU\ count\ from\ the\ July\ 2007\ ACS\ MAF\ extracts}{HU\ count\ from\ the\ 2007\ PEP\ HU\ estimates}$$

This added records from counties with a small percentage of non city-style addresses and from fast growing counties with more than 10 percent undercoverage. For counties with a small number of non city-style addresses there should be little chance of duplication by using the ungeocoded addresses. Ungeocoded records from 63 counties (see Table 6 for a list of counties) were added to the 2009 frame as a result.

In late July / early August of 2009, a quick decision was needed on whether to continue allowing these records to be in the ACS universe or not. If the new records result in interviews, then we want to continue including them. The main point of the research was not to make statistical

[3] The process by which ACS edits and prepares the MAF extracts to create the ACS sampling universe.

comparisons, but rather to determine if there was enough evidence to continue using the ungeocoded records in these 63 counties.

To assess whether the added records result in a high percentage of interviews, we decided to investigate the mail, Computer Assisted Telephone Interviewing (CATI) and Computer Assisted Personal Interviewing (CAPI) data collection modes, and interview status of the records.

3. Methodology

We began by investigating all of the new ungeocoded records, both mailable and unmailable[4], that were used in the sample for the January – May 2009 panels and included in the ACS control file. We restricted the records to the January through May panels because, at the time of the analysis, not all of the data was available for the panels started after May.

First we defined the study universes:

- Total Sample (Added): These are the ungeocoded sample records from the 63 counties where we began using ungeocoded records from the January – May 2009 ACS panels
- Total Sample (Other Ungeocoded): These are all of the ungeocoded sample records from the January – May 2009 ACS panels that were not in the 63 counties

The following tallies were generated for both universes listed above:

- Mail returns – cases where an ACS mail questionnaire was returned
- Mailable sample – cases that ACS was able to mail to
- CATI interviews – cases where ACS obtained a telephone interview
- CAPI interviews – cases where ACS obtained an interview through a personal visit
- Records sampled out of CAPI – cases that were not included in the subsample of CAPI eligible cases
- CAPI noninterviews (NIs) – cases where ACS did not obtain an interview through a personal visit
- CAPI Deletes – cases that were discovered to be ineligible for CAPI upon a visit; e.g. structures under construction, unoccupied mobile home/trailer/tent sites, demolished buildings, etc.

The tallies were generated on an overall basis as well as by county. From these tallies we computed five ratios from both universes:

[4] There were only 2 unmailable records among the records added from the 63 counties

- Weighted Total Interview ratio[5] = $\frac{mail\ returns + CATI\ interviews + CAPI\ interviews}{Total\ Sample}$
- Mail check-in ratio = $\frac{mail\ returns}{mailable\ sample}$
- CAPI interview ratio = $\frac{CAPI\ interviews}{CAPI\ interviews + CAPI\ deletes + CAPI\ NIs}$
- CAPI NI ratio = $\frac{CAPI\ NIs}{CAPI\ interviews + CAPI\ deletes + CAPI\ NIs}$
- CAPI delete ratio = $\frac{CAPI\ deletes}{CAPI\ interviews + CAPI\ deletes + CAPI\ NIs}$

We calculated a weighted total interview ratio because of different initial and CAPI sampling rates. We used the same weighting methodology as quality measures to calculate the weights.

While we used the CATI cases in calculating the weighted total interview ratio, we decided not to directly compare the CATI cases between the two universes, as our primary concern was whether we could mail to the added cases or not. We did directly compare the CAPI cases, as they were able to give insight as to how comparable the added units are (i.e., how often they are deleted, or found to be non-existent).

We also investigated the number of units that were Undeliverable as Addressed (UAA) and the UAA ratio for the records by county for the mailable records from Total Sample (Added) and from Total Sample (Other Ungeocoded). The UAA ratio is defined as:

- UAA ratio = $\frac{UAAs}{mailable\ sample}$

A record was defined as UAA if the address came back as UAA for both the first and second mailings.

4. Results

4.1 Did we get sufficient interviews for the newly added ungeocoded units, and what type of interview were they?

We compared ratios for the newly added ungeocoded records in the 63 counties to the ungeocoded records in the other counties. (See Table 1)

When we compare the ratios, we find that most of the rates are fairly similar. The mail check-in ratio for the added ungeocoded is almost five percentage points higher than for the Other Ungeocoded.

We are getting a good total interview ratio for the added records (79.3 percent).

[5] Weights account for sampling probabilities and CAPI subsampling

Table 1. Ratios for the Total Sample (Other Ungeocoded) and Total Sample (Added) records

Ratios	Total Sample (Other Ungeocoded)	Total Sample (Added)[1]
Mail check-in* (mail returns / sample)	39.2%	44.2%
Undeliverable as Addressed* (UAA / sample)	20.7%	20.4%
CAPI Interview* (CAPI interviews / sent to CAPI[6])	60.5%	63.6%
CAPI Deletes* (CAPI deletes / sent to CAPI)	35.0%	34.9%
CAPI Non-Interview* (CAPI non-interviews / sent to CAPI)	4.4%	1.4%
Weighted Total Interview	76.6%	79.3%

Source: U.S. Census Bureau, 2009 American Community Survey January – July 2009
*these are unweighted ratios
1. The Addded only had 2 unmailable addresses

Undeliverable as Addressed

There were 288 UAAs in Total Sample (Added). Bear in mind that while a UAA record would not return a mail interview, it could be interviewed in CATI or CAPI instead. There were 96 UAA addresses that remained after CAPI subsampling. Interviews were obtained for 37 Total Sample (Added) units in CAPI. The remaining records were deleted by CAPI. (Table 2)

Table 2. Interview status of UAAs

Interview Status	Total Sample (Other Ungeocoded)	Total Sample (Added)
CAPI interview	1,707 30.3%	37 38.5%
CAPI delete	3,901 69.2%	59 61.4%
CAPI NI	28 0.5%	0 0%

Source: U.S. Census Bureau, 2009 American Community Survey January – July 2009

[6] Sent to CAPI includes the records that remained in the CAPI universe after subsampling.

4.2 Did we get many type C non-interviews?

CAPI Deletes

The CAPI delete ratio for Total Sample Added (34.9 percent) is close to the ratio for the Total Sample Other Ungeocoded (35.0 percent). (Table 1) Taking a closer look at the CAPI deletes, we examined the CAPI final outcome codes for them. We suspected that a large number of the CAPI deletes would fall under 'under construction'; this was not the case. Table 4 shows that almost 60 percent of the CAPI deletes fell under the category 'address nonexistent'.

For the most part, the ratios for the Total Sample Added records were close to the Total Sample Other Ungeocoded with the notable exceptions being the "Other" and "Address nonexistent" categories. (Table 3)

Table 3. CAPI final outcomes for CAPI deletes

CAPI final outcome	Total Sample Other Ungeocoded		Total Sample Added	
	Frequency	Percent	Frequency	Percent
Under Construction	184	3.4	1	1.4
Demolished	176	3.2	1	1.4
House/trailer moved	153	2.8	3	4.1
Converted to permanent business or storage	232	4.3	2	2.7
Other	2,654	48.7	23	31.5
Address nonexistent	1,753	32.2	43	58.9
Unit nonexistent – Basic street address found	131	2.4	--	--
Group Quarters	148	2.7	--	--
Merged with another unit	7	0.1	--	--
Condemned	7	0.1	--	--

Source: U.S. Census Bureau, 2009 American Community Survey January – July 2009

4.3 What is the likelihood of duplication?

To assess the likelihood of whether the ungeocoded units were duplicates of existing units or not, we investigated the ACT codes for the Total Sample Added records for which a successful interview (mail, CAPI, or CATI) was obtained. Of the 867 interviewed records, 415 received a geocode and thus had an ACT code. Table 5 shows that most of these records (97.8 percent) come from blocks where we would use new geocoded DSF records that are outside of the blueline[7]. This indicates that of the ungeocoded records that have an

[7] This is where there are more than 80 percent city-style addresses in a block.

ACT code, most would be included if the criteria for new geocoded records were applied. Also, taking into consideration that about 95 percent of these records are in a block where at least 90 percent of units have a city-style address, the likelihood of duplication is small.

Table 4. ACT for records that had an interview (mail, CATI, or CAPI)

ACT	Frequency	Percent
100% city-style addresses	187	45.0
95%-99.99% city-style addresses, some have a DSF source	177	42.7
90%-94.99% city-style addresses, some have a DSF source	32	7.7
85%-89.99% city-style addresses, some have a DSF source	8	1.9
80%-84.99% city-style addresses, some have a DSF source	2	0.5
75%-&9.99% city-style addresses, some have a DSF source	2	0.5
0.01%-69.99% city-style addresses, some have a DSF source	1	0.2
No addresses	6	1.5

Source: U.S. Census Bureau, 2009 American Community Survey January – July 2009

4.4 Are there areas that should not be included?

In considering if there were areas that should not have been included, we studied the nine counties where the weighted total interview ratio was less than 50 percent. These counties had a small sample size (four had15-17 cases and five had four or fewer cases from Jan-May 2009). Interview ratios for counties with sample sizes this small may not be reliable. With such little data at a county level and positive overall results, we decided to continue to use the ungeocoded records in all counties.

4.5 Did the results hold up over all of 2009?

We decided to investigate these ratios again, using all of the 2009 panels, to determine if our initial findings remained true throughout the rest of the year. We found that the ratios remained similar when calculated for all 2009 panels. (Table 5) This result reinforces the decision to continue including these records in the frame.

Table 5. Ratios for records from the January – December 2009 (All) and January – December 2009 (Added)

Ratios	Total Sample Other Ungeocoded January – December 2009	Total Sample Added January – December 2009
Mail Return	38.6%	44.2 %
UAA	21.0%	20.0 %
CAPI Interview	61.4%	66.9 %
CAPI Deletes	34.5%	30.8 %
CAPI Non-Interview	4.0%	2.3 %
Weighted Total Interview	78.1%	81.4%

Source: U.S. Census Bureau, 2009 American Community Survey January 2009 – February 2010

We also revisited the question of whether there were areas that we should not have included. When we take the whole year of sampling into consideration, most of the areas we previously noted no longer had the low interview ratio they had during the January-May panels. Based on this information, excluding records from these types of areas this is a minimal concern.

5. Conclusion

As a result of the favorable weighted total interview ratio (79.3 percent) from the ungeocoded units in the 63 counties, we decided to continue using them in the ACS Frame.

Table 6. State/County Codes

State/County Code	State	County
06009	California	Calaveras County
06043	California	Mariposa County
08007	Colorado	Archuleta County
08015	Colorado	Chaffee County
08037	Colorado	Eagle County
08039	Colorado	Elbert County
08067	Colorado	La Plata County
08091	Colorado	Ouray County
08097	Colorado	Pitkin County
08107	Colorado	Routt County
08117	Colorado	Summit County
12119	Florida	Sumter County
12129	Florida	Wakulla County
13013	Georgia	Barrow County
13035	Georgia	Butts County
13085	Georgia	Dawson County
13103	Georgia	Effingham County
13133	Georgia	Greene County
13137	Georgia	Habersham County
13157	Georgia	Jackson County
13159	Georgia	Jasper County
13171	Georgia	Lamar County
13177	Georgia	Lee County
13211	Georgia	Morgan County
13219	Georgia	Oconee County
13223	Georgia	Paulding County
13231	Georgia	Pike County
13233	Georgia	Polk County
13241	Georgia	Rabun County
13291	Georgia	Union County
13297	Georgia	Walton County
13311	Georgia	White County
20069	Kansas	Gray County
21103	Kentucky	Henry County
21215	Kentucky	Spencer County
22089	Louisiana	St. Charles Parish
22093	Louisiana	St. James Parish
26009	Michigan	Antrim County
26019	Michigan	Benzie County
26101	Michigan	Manistee County

27019	Minnesota	Carver County
27031	Minnesota	Cook County
27147	Minnesota	Steele County
28137	Mississippi	Tate County
32019	Nevada	Lyon County
37019	North Carolina	Brunswick County
37029	North Carolina	Camden County
37145	North Carolina	Person County
37189	North Carolina	Watauga County
46057	South Dakota	Hamlin County
46087	South Dakota	McCook County
48425	Texas	Somervell County
49029	Utah	Morgan County
51043	Virginia	Clarke County
51061	Virginia	Fauquier County
51073	Virginia	Gloucester County
51101	Virginia	King William County
51137	Virginia	Orange County
55041	Wisconsin	Forest County
55113	Wisconsin	Sawyer County
55125	Wisconsin	Vilas County
56023	Wyoming	Lincoln County