



April 30, 2014

2014 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT
MEMORANDUM SERIES #ACS14-RER-15

DSSD 2013 AMERICAN COMMUNITY SURVEY MEMORANDUM SERIES #ACS13-UC-05

MEMORANDUM FOR ACS Research and Evaluation Advisory Group

From: Patrick J. Cantwell/*Signed*/
Division Chief, Decennial Statistical Studies Division

Prepared by: Dianne Aubuchon
Community Address Updating Systems Branch
Decennial Statistical Studies Division

Subject: Assessment of the Revised Duplication Zone

Attached is the final American Community Survey Research and Evaluation report for the Assessment of the Revised Duplication Zone. The revised duplication zone allows for the use of ungeocoded United States Postal Service Delivery Sequence File records in counties with at least 80 percent city-style addresses. The goal of this evaluation is to determine whether this rule has a negative effect on American Community Survey interview and delete rates.

If you have any questions about this report, please contact Dianne Aubuchon (301) 763-7706 or Tony Tersine (301) 763-1994.

Attachment

cc:

ACS Research and Evaluation Work Group

Eric Slud (CSRM)

Jeanne Woodward (SEHSD)

Steve Hefter (DSSD)

Tony Tersine

Beth Tyszka

Assessment of the Revised Duplication Zone

FINAL REPORT



Dianne Aubuchon
Decennial Statistical Studies Division

Executive Summary

The 2012 American Community Survey Main Phase Housing Unit Frame included ungeocoded¹ records from the United States Postal Service Delivery Sequence File for 730 counties where ungeocoded records were previously excluded from the frame. These 730 counties contained at least 80 percent city-style addresses². To assess this change we looked at the total interview rate and type C³ non-interview rate and how they changed by using these records.

In the 2,135 counties where we used ungeocoded Delivery Sequence File records in both 2011 and 2012 we found a difference of 0.15 percentage points for the total interview rate and a difference of 0.17 percentage points for the type C non-interview rate when including the ungeocoded Delivery Sequence File records. In the 730 counties where we started using ungeocoded Delivery Sequence File records in 2012, the difference in total interview rate was 0.18 percentage points and the difference in type C non-interview rate was 0.19 percentage points. All of these differences were within predefined acceptable limits.

Since all differences were within acceptable limits, we recommend that we continue to use ungeocoded Delivery Sequence File records in counties with at least 80 percent city-style addresses.

1. Introduction

The Master Address File/Topologically Integrated Geographic Encoding and Referencing database (MTdb) contains addresses for all known living quarters and certain nonresidential units in the U.S. and Puerto Rico. The MTdb is the sole source of housing unit (HU) addresses on the American Community Survey (ACS) HU sampling frame. The largest source of address updates to the MTdb between censuses is the Delivery Sequence File⁴ (DSF) from the U.S. Postal Service (USPS). The DSF is a file of mail delivery points serviced by the USPS. An updated DSF is used to update the MTdb every six months.

Not all new DSF addresses are included on our sampling frame. The ACS has historically excluded DSF addresses in areas where those addresses are more likely to duplicate existing addresses that do not have a city-style address. Those areas with higher duplication risk are referred to as being “in the duplication zone.” The duplication zone is mainly based on the percentage of city-style addresses in a given area.

¹ Records that do not have any block information are classified as ungeocoded.

² An address is city-style if it has a house number and street name; otherwise, it is non city-style.

³ A non-interview resulting from 1) a unit no longer qualifying as a housing unit or 2) a unit no longer existing.

⁴ A file containing all mailing addresses serviced by the USPS.

There is less concern over the addresses in areas with a high percentage of city-style addresses because address matching can be done for city-style addresses and is relatively accurate. Address matching cannot match a city-style address to a noncity-style address.

Since the 2000 Census, the ACS has allowed the use of geocoded DSF units in areas that are within the 2000 mailout/mailback areas. 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. For the 2005 ACS Frame, ACS revised the duplication zone and began using the Address Characteristic Type (ACT) (block level codes that describe the address types found in the block) code to expand where we allow the use of geocoded DSF records.

Prior to 2012, ACS used the type of enumeration area from the 2000 Census to determine where to include ungeocoded DSF addresses. Records that were either completely or partially within the 2000 mailout/mailback areas were used. Ungeocoded DSF records are not assigned to a block, so the alteration in 2005 to use ACT codes did not affect where we used the ungeocoded DSF records. Because the duplication zone for ungeocoded DSF records was still based upon information from the 2000 Census, we needed to come up with a new definition.

We wanted to use a definition similar to that used for geocoded records. Previous research indicated that the risk of duplication was low enough when a cut-off of 80 percent city-style addresses was used (Aubuchon, 2012). Therefore, beginning in 2012, ungeocoded DSF records in counties with 80 percent or more city-style addresses were included in the ACS sampling frame. This change resulted in ACS including ungeocoded addresses in 730 counties where they were previously excluded.

The purpose of this research is to assess the revised duplication zone to determine if this new definition has had an adverse affect on ACS interview and delete rates.

2. Methodology

In order to assess the revised duplication zone we will look at two sets of counties within the 2012 sample. First, we will look at the counties where we used ungeocoded DSF records in both 2011 and 2012. Second, we will look at the 730 counties where we began using ungeocoded DSF records. We will look at rates for the following universes:

- All sample records
- All sample records except ungeocoded post Census⁵ DSF adds
- All post Census DSF adds
- All geocoded post Census DSF adds

⁵ We are referring to records that were added post 2010 Census

- All ungeocoded post Census DSF adds

We will calculate a type C non-interview rate and total interview rate for these sub-universes of the 2012 sample.

Type C non-interviews result from some Computer Assisted Personal Interviewing (CAPI) cases. This includes cases that were marked as under construction, demolished, a vacant trailer site, a permanent business or storage, merged with another unit, condemned, unit non-existent but the basic street address was found, non-existent address, or group quarters.

The type C non-interview rate is the percentage of all sample cases that result in a type C non-interview in CAPI:

$$\text{Type C Non-Interview Rate} = \frac{\text{Type C Non-interviews}}{\text{Total Sample}} \times 100$$

A high type C non-interview rate indicates that we are including records on the frame that should be excluded. The type C non-interview rate will be weighted using the replicate weights that take into account CAPI subsampling.

The total interview rate is the percentage of all sample cases that resulted in an interview from mail, Computer Assisted Telephone Interviewing (CATI), or CAPI:

$$\text{Total Interview Rate} = \frac{\text{Mail Returns} + \text{CATI Interviews} + \text{CAPI Interviews}}{\text{Total Sample}} \times 100$$

The total interview rate will be weighted using the CAPI subsampling weights as well.

For both sets of counties, we will use a t-test for correlated samples to compare the two rates between the first two universes listed above. We will be taking the difference between the estimated rates, standardizing by the square root of the weighted variance estimator of that difference, and comparing the absolute value to standard-normal cutoffs. We will be using the successive differences replication (SDR) method of variance estimation from the ACS Design and Methodology (U.S. Census Bureau, 2010). The following formulas were used:

- Standard Error = $\sqrt{\frac{4}{80} \sum_1^{80} (\text{Difference}_i - \text{Difference}_0)^2}$
- T-statistic = $\frac{\text{Difference}_0 - \text{Null Value}}{\text{Standard Error}}$

Where $\text{Difference}_i = \text{Rate1}_i - \text{Rate2}_i$

We are willing to accept a decrease of as much as 0.2 percentage points in the total interview rate and up to a 0.5 percentage point increase in the type C non-interview rate. The other universes are for informational purposes only. All of the testing done for this report used a 90% confidence level.

We will also calculate the following five rates for informational purposes only:

- The mail check-in rate – the percentage of sample records, that ACS mails a form to, that return an acceptably complete form

$$\text{Mail Check-In Rate} = \frac{\text{Mail Returns}}{\text{Mailable Sample}} \times 100$$

- The CAPI interview rate – the percentage of sample records for which a field representative (FR) was sent out to interview, that an interview was obtained

$$\text{CAPI Interview Rate} = \frac{\text{CAPI Interviews}}{\text{All CAPI Cases}} \times 100$$

- The CAPI non-interview rate – the percentage of sample records for which an FR was sent out to interview and was unable to obtain an interview. Possible reasons for a non-interview include a language barrier with the interviewee or that no one was home during the interview period.

This rate includes the type C non-interviews, which mark a unit as no longer existing or being an HU. It also includes other types of non-interviews where units are still existing HUs but an eligible or cooperative respondent could not be found for interview.

$$\text{CAPI Non-Interview Rate} = \frac{\text{CAPI Non-Interviews}}{\text{All CAPI Cases}} \times 100$$

- The frame delete rate – the percentage of all sample cases that result in a frame delete. Frame deletes are CAPI type C non-interviews that result in the HU being deleted from all future ACS sampling frames.

$$\text{Frame Delete Rate} = \frac{\text{Cases Deleted from the Frame}}{\text{Total Sample}} \times 100$$

- The Undeliverable as Addressed (UAA) rate – the percentage of sample records where a survey that was mailed out was returned by the USPS as undeliverable.

$$\text{Undeliverable as Addressed Rate} = \frac{\text{Cases Returned as Unmailable}}{\text{Mailable Sample}} \times 100$$

These five rates will be weighted similarly to the non-interview and total interview rates above.

If we see an increase of more than 0.5 percentage points in the non-interview rate or a decrease of more than 0.2 percentage points in the total interview rate, then we will calculate the same rates for alternate duplication zone definitions based on higher percentages of city-style addresses in the county. We will use the same methodology to compare rates as mentioned above.

3. **Limitations**

As we are looking at overall sample rates, there could be some large localized effects due to these changes.

The rates calculated here are not final rates. The data that were used to calculate the rates have not gone through all of the steps necessary to determine the final status. As a result of those steps, the status of a case could change.

4. **Results**

4.1 Counties Where We Used Ungeocoded DSF Records in both 2011 and 2012

As expected, the ungeocoded DSF adds have a lower interview rate and a higher type C non-interview than the geocoded DSF adds. Adding ungeocoded DSF records to the frame, therefore, results, in a lower overall interview rate and higher type C non-interview rate for the sample.

Looking at the total interview rates in Table 1 below, we found the rate for all sample records to be 88.8 percent and the rate for the sample without ungeocoded DSF adds to be 89.0 percent. We found the difference between these two rates to be 0.15 percentage points. Our null hypothesis was that the difference in total interview rates was less than or equal to 0.2 percentage points. We calculated a t-statistic of -9.98 for this test, and concluded that we did not exceed our limit of a 0.2 percentage point decrease in the total interview rate.

Next, we looked at the type C non-interview rates in Table 1. We found a rate of 2.5 percent for all sample records, and a rate of 2.3 percent for all sample records except for the ungeocoded post Census DSF adds. We found a difference of 0.17

percentage points between these two rates. Our null hypothesis was that the difference in type C non-interview rates was less than or equal to 0.5 percentage points. We calculated a t-statistic of -59.3 for this test, and concluded that we did not exceed our limit of a 0.5 percentage point increase in the type C non-interview rate.

When looking at just the post census DSF adds, we see that the total interview rate is much lower than the overall rate and that the type C non-interview rate is much higher.

Table 1. Total Interview and Type C Non-Interview Rates for the 2,135 Counties Where the Duplication Zone Status Did Not Change

	Total Interview Rate (Standard Error)	Type C Non-Interview Rate (Standard Error)
2012 Sample	88.8% (0.0002)	2.5% (0.0002)
2012 Sample Without Ungeocoded DSF Adds	89.0% (0.0002)	2.3% (0.0001)
DSF Adds	68.3% (0.0047)	25.2% (0.0047)
Geocoded DSF Adds	71.3% (0.0054)	21.3% (0.0052)
Ungeocoded DSF Adds	64.9% (0.0071)	29.5% (0.0073)

Source: U.S. Census Bureau, 2012 American Community Survey

We looked at several other rates as well, seen in Table 2 below. The mail check-in and CAPI interview rates are both statistically significantly lower when looking solely at the post census DSF adds. The frame delete and UAA rates on the other hand, are statistically significantly higher when looking at the post census DSF adds.

Table 2. Miscellaneous Rates for Counties Where the Duplication Zone Status Did Not Change

	Mail Check-In Rate (Standard Error)	Undeliverable as Addressed Rate (Standard Error)	CAPI interview Rate (Standard Error)	CAPI non-interview Rate (Standard Error)	Frame Delete Rate (Standard Error)
2012 Sample	42.9% (0.0004)	10.7% (0.0003)	88.9% (0.0004)	6.0% (0.0004)	2.7% (0.0001)
2012 Sample Without Ungeocoded DSF Adds	43.0% (0.0004)	10.6% (0.0003)	89.1% (0.0004)	6.0% (0.0003)	2.5% (0.00060001)
DSF Adds	33.1% (0.0037)	25.8% (0.0043)	55.1% (0.0075)	3.7% (0.0024)	21.0% (0.0033)
Geocoded DSF Adds	34.2% (0.0053)	22.2% (0.0051)	59.9% (0.0088)	4.5% (0.0034)	21.5% (0.0042)
Ungeocoded DSF adds	31.9% (0.0056)	30.0% (0.0069)	50.0% (0.0110)	2.8% (0.0031)	20.5% (0.0052)

Source: U.S. Census Bureau, 2012 American Community Survey

4.2 Counties Where We Began Using Ungeocoded DSF Records in 2012

Looking at the total interview rates in Table 3, we found the rate for all sample records to be 87.3 percent and the rate for all sample records except the ungeocoded post census DSF adds to be 87.5 percent. We found the difference between these two rates to be 0.18 percentage points. Our null hypothesis was that the difference in total interview rates was less than or equal to 0.2 percentage points. We calculated a t-statistic of -0.49 for this test, and concluded that we did not exceed our limit of a 0.2 percentage point decrease in the total interview rate.

Next, we looked at the type C non-interview rates in Table 3. We found a rate of 5.9 percent for all sample records, and a rate of 5.7 percent for all sample records except the ungeocoded post Census DSF adds. We found a difference of 0.19 percentage points between these two rates. Our null hypothesis was that the difference in type C non-interview rates was less than or equal to 0.5 percentage points. We calculated a t-statistic of -10.4 for this test, and concluded that we did not exceed our limit of a 0.5 percentage point increase in the type C non-interview rate.

When looking at just the Post Census DSF adds, we see that the total interview rate is much lower than the overall rate and that the type C non-interview rate is much higher. However, when comparing the geocoded and ungeocoded DSF adds, the rates were not significantly different from each other.

Table 3. Total Interview and Type C Non-Interview Rates for Counties that Moved Outside the Duplication Zone

	Total Interview Rate (Standard Error)	Type C Non-Interview Rate (Standard Error)
All Sample Records	87.3% (0.001)	5.9% (0.001)
All Sample Records Except Ungeocoded DSF Adds	87.5% (0.001)	5.7% (0.001)
All DSF Adds	66.9% (0.016)	26.6% (0.015)
All Geocoded DSF Adds	66.8% (0.023)	26.2% (0.0)
All Ungeocoded DSF Adds	66.9% (0.024)	27.0% (0.023)

Source: U.S. Census Bureau, 2012 American Community Survey

We looked at several other rates as well, seen in Table 4 below. The mail check-in rate and CAPI interview rate have a statistically significant drop when looking at just the post Census DSF adds, much like the total interview rate in Table 3. The frame delete rate has a statistically significant increase when looking at just the post census DSF adds.

Table 4. Miscellaneous Rates for Counties that Moved Outside the Duplication Zone

	Mail Check-In Rate (Standard Error)	Undeliverable as Addressed Rate (Standard Error)	CAPI interview Rate (Standard Error)	CAPI non-interview Rate (Standard Error)	Frame Delete Rate (Standard Error)
All Sample Records	33.0% (0.002)	30.8% (0.002)	86.8% (0.002)	3.1% (0.001)	4.4% (0.001)
All Sample Records Except Ungeocoded DSF Adds	33.1% (0.001)	30.8% (0.002)	87.1% (0.002)	3.2% (0.001)	4.1% (0.001)
All DSF Adds	29.0% (0.014)	27.6% (0.016)	55.9% (0.022)	2.2% (0.007)	29.2% (0.017)
All Geocoded DSF Adds	31.7% (0.019)	24.7% (0.020)	54.0% (0.033)	2.8% (0.012)	31.2% (0.021)
All Ungeocoded DSF adds	26.8% (0.021)	30.2% (0.024)	57.4% (0.031)	1.8% (0.008)	27.6% (0.027)

Source: U.S. Census Bureau, 2012 American Community Survey

5. Summary

When including the ungeocoded post census DSF adds in our universe, we did see the total interview rate decrease and the type C non-interview rate increase. However, we found the difference in total interview and type C non-interview rates for counties where we continued using ungeocoded DSF records and counties where we started using ungeocoded DSF records for the first time in 2012 to be within acceptable limits.

We recommend that we continue to use ungeocoded DSF records in counties with at least 80 percent city-style addresses.

6. References

Aubuchon, Dianne. (2012), "Use of Ungeocoded Delivery Sequence File Records in the American Community Survey Sampling Frame," 2012 American Community Survey Research And Evaluation Report Memorandum Series #ACS12-RER-21, May 4, 2012.

U.S. CENSUS BUREAU. (2010), American Community Survey Design and Methodology Chapter 12, December 2010.

http://www.census.gov/acs/www/Downloads/survey_methodology/Chapter_12_RevisedDec2010.pdf