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

February 3, 2012

2012 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT MEMORANDUM SERIES #ACS12-RER-03

DSSD 2012 American Community Survey Research Memorandum Series ACS12-R-01

MEMORANDUM FOR ACS Research and Evaluation Steering Committee

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

Prepared by: Larry Bates
Jim Hartman
Community Address Updating Systems Branch
Decennial Statistical Studies Division

Subject: Research on Master Address File Quality – Implications for the American Community Survey

Attached is the final American Community Survey Research and Evaluation report for Research on Master Address File (MAF) Quality – Implications for the American Community Survey (ACS). This research was undertaken prior to the 2010 Address Canvassing to summarize research that had been done to assess the quality of the MAF and to identify components of the ACS sample frame requiring additional research.

If you have any questions about this report, please contact Jim Hartman (301-763-1976) or Larry Bates (301-763-5926).

Attachment

cc:
ACS Research and Evaluation Team
Matt Zimolzak (ACSO)
Cliff Loudermilk (DSMD)
Bob Colosi (DSSD)
Rodger Johnson (POP)
Jason Devine (POP)

February 3, 2012

Research on Master Address File Quality – Implications for the American Community Survey

Larry Bates
Jim Hartman
Decennial Statistical Studies Division

United States™ Census Bureau

Research on Master Address File Quality - Implications for the American Community Survey

1 Executive Summary

There appears to be both overcoverage and undercoverage in the American Community Survey housing unit frame. Overall, American Community Survey housing unit counts are consistently higher than housing unit estimates and various measures of "truth." National gross overcoverage was estimated at 10.2 percent and gross undercoverage at 6.4 percent. The largest component of undercoverage is units that are missing from the Master Address File altogether.

Undercoverage in rural areas, particularly those areas experiencing growth, is an issue. The Frame Assessment for Current Household Surveys National Evaluation estimated the gross undercoverage in current survey area frame blocks to be near 50 percent. The growth in noncity-style addresses on the Delivery Sequence File may be able to be used to target areas of high growth in rural areas. More research on this is necessary.

Undercoverage of new construction exists outside of rural areas as well. The Delivery Sequence File eventually includes much of the new construction, but processing lags from the time a new construction address appears in the Delivery Sequence File until the time the American Community Survey uses the address results in undercoverage.

Mobile homes also have a high undercoverage rate, regardless of whether they are located in rural areas or other areas.

Overcoverage is likely due to the inclusion of certain categories of Delivery Sequence File records; but removing whole classes of Delivery Sequence File records could increase the already high undercoverage rate of new construction addresses. The use of various classes of Delivery Sequence File addresses (particularly Excluded from Delivery Statistics records) merits further research.

Further research into the definition of the "duplication zone" (areas where Delivery Sequence File records are not used because of an unacceptable risk of duplication with existing addresses) needs to be done to identify areas where the American Community Survey should expand or discontinue the use of new Delivery Sequence File addresses. County-level growth rates should be included in the definition of the duplication zone.

Results from the 2010 Census address canvassing operation should be used to assess various categories of addresses. Using the address canvassing results as ground truth could allow us to determine if there are types of records in the American Community Survey housing unit frame that are more likely to represent overcoverage than others. At the same time we can examine the

addresses added by address canvassing that matched to other records in the Master Address File to determine if there are categories of addresses in the Master Address File being excluded by the filter that we should be including.

2 Introduction

The Master Address File (MAF) is a database containing addresses for all known living quarters in the United States. It serves as the sole source of housing unit addresses for the American Community Survey (ACS) sampling frame. The MAF is updated twice a year with delivery point addresses from the United States Postal Service (USPS) Delivery Sequence File (DSF). Updates from various field operations conducted by the Census Bureau are also applied to the MAF. It is the best national source of addresses, updated regularly, available to the ACS for use as a sampling frame.

Various areas throughout the Census Bureau have conducted research on the quality of the MAF. The Decennial Statistical Studies Division (DSSD) computes national estimates of coverage of the MAF every year as part of the Address Coverage Improvement and Evaluation Program (ACIEP). DSSD also looked at the Demographic Area Address Listing (DAAL) program and certain subcategories of DSF records.

The Demographic Statistical Methods Division (DSMD) analyzed permit new construction and the current surveys' area frame through the Frame Assessment for Current Household Surveys (FACHS) program. The area frame covers areas where permits are not issued for new construction and/or areas where at least four percent of the addresses have an incomplete city-style address (i.e., lack a house number, street name, or both). These are mostly rural areas and include areas where the DSF is not currently used to update the MAF. The goal of the FACHS program is to determine if the MAF can be used to replace the multiple frames used by current surveys without compromising coverage and quality. As part of this program the DSMD selected a national sample of blocks to be listed so that they could produce coverage estimates. Results from these block listings were also used to evaluate filter rules that could be used to determine which addresses should be included in a MAF-based sampling frame.

The Population Division (POP) researched potential methods for using the MAF to assist them in calculating population estimates. This involved comparing housing unit counts from the ACS sampling frame to housing unit estimates. They also looked at alternative filters that could potentially be used to help calculate housing unit estimates using the MAF. POP also compared the change in the housing unit estimates to the change in the ACS housing unit counts as part of an effort to better understand the differences between the housing unit estimates and the ACS housing unit frame.

While the various individual research projects had differing objectives, together they provide a general idea of areas where the MAF may be deficient and where the ACS sampling frame could be improved. The goal of this report is to determine what the current research says about the

quality of the ACS sampling frame – both in terms of overcoverage and undercoverage. Overcoverage impacts the ACS by causing otherwise ineligible units to be eligible for sample. ACS must send interviewers to determine the units are ineligible using valuable resources. Undercoverage impacts the ACS by not giving eligible units a chance of selection. While the weighting process for the ACS adjusts for the fact that these units were not eligible, their characteristics may or may not be the same as the units that will represent them in the ACS estimates.

In addition, we want to find out if there are categories of address records in the MAF that have either not been investigated or that could use further research. The goal of this project is to examine the research that has already been done, assess the implications for the ACS, and determine what more could be done.

3 Results

Since this project was a review of existing research, this section of the report provides a brief summary of the research and how the findings impact the ACS. For more detailed analysis of the findings in each report, please see the individual reports listed in the References section at the end of this report.

There were four research questions we attempted to answer as part of this review.

3.1 Research Question #1. What does the previous research indicate about the quality of coverage for the ACS?

3.1.1 National Coverage

Several reports show that there is net overcoverage in the ACS sampling frame. The National Estimates of Coverage report estimates that this net overcoverage has increased from 2.0 percent in 2002 to 4.1 percent in 2007 at the national level (Johnson, 2008). POP research has also shown that the ACS housing unit counts are growing more rapidly and diverging from the housing unit estimates (Devine et al., 2008).

The net overcoverage may be somewhat overstated because of the inconsistent reference dates between the various reports and the ACS. The reference date for the National Estimates of Coverage report is July 1 of the current year but the MAF extracts used to compute these estimates are used by the ACS to select the sample for the following sample year. In the 6 to 18 months between the time the ACS receives the MAF extracts and the time that the ACS contacts a given sample unit, there will be some number of housing unit records in the MAF extract that represent new completed construction. While there is substantial evidence that there is net overcoverage in the ACS sampling frame, the amount of overcoverage is overstated somewhat because of the inconsistent reference dates.

The FACHS National Evaluation produced estimates of gross overcoverage and gross undercoverage on the MAF. Gross overcoverage at the national level was measured at 10.2 percent while gross undercoverage was 6.4 percent. The gross undercoverage includes a 4.7 percent omission rate and 1.7 percent erroneous exclusions (Li, 2008). Erroneous exclusions are records that appear on the MAF but are excluded from the ACS sampling frame. Erroneous exclusions could be reduced by modifying the filter as long as a filter rule can be defined that includes those records without adding a lot of additional noise to the frame. Omissions on the other hand are more problematic in that those units are missing from the MAF altogether. Until those omissions are added to the MAF they will continue to be missed by the ACS.

The overcoverage is a concern. While the DAAL program is adding units to the MAF, most of the growth in the ACS housing unit counts is coming from the DSF. Given the undercoverage of new construction, which is discussed further below, it seems that many of the records added to the ACS housing unit frame are records that perhaps should be excluded. Various research (Johnson, 2008 and Devine et al., 2008) has identified both Excluded from Delivery Statistics (EDS) records[1] and ungeocoded records[2] from the DSF as likely sources of this overcoverage. Both of these categories of addresses will be discussed in a later section.

3.1.2 Regional Coverage

The research shows that coverage tends to be worse in the South than in any other region. The 2007 National Estimates of Coverage Report estimates the net overcoverage in the South to be 5.24 percent. Reese also determined that the ACS counts and the housing unit estimates are diverging the most in the South with the ACS counts growing more rapidly than the housing unit (HU) estimates.

The FACHS Area Frame Study and the FACHS National Evaluation both indicate that gross overcoverage is highest in the South. The overcoverage can most likely be attributed to the high growth rate in the South. Gross undercoverage, omissions, and erroneous exclusions are also highest in the South due to the larger amount of rural areas in that part of the country and the problems associated with capturing new addresses in those rural areas.

3.1.3 County-Level Coverage

Reese compared the change in the ACS counts to the change in the HU estimates and found that the two are diverging. He found that, for smaller counties, there was more change in the housing unit estimates. This is to be expected since the smaller counties are more likely to be rural counties where the ACS does not utilize many of the records on the DSF. The ACS therefore is not picking up growth in those areas. For the larger counties there was more change in the ACS counts, probably because those are the counties with the most growth.

[1] EDS records are discussed in section 3.3.1
[2] Ungeocoded records are discussed in section 3.2.1

Devine, et al. compared the county-level ACS housing unit counts to housing unit estimates for various size and growth rate categories as part of an effort to determine if the MAF can be used to improve the county-level housing unit estimates process. They found that the ACS counts were higher than the housing unit estimates for all county size and growth rate categories except for small counties with moderate to high growth between 2000 and 2006.

Of particular concern is the 7.7 percent difference for the small counties (less than 10,000 housing units) with a high growth rate (15 percent or more). These smaller counties are probably the more rural counties where both omissions and erroneous exclusions are a problem. If the new housing units in these counties have noncity-style addresses then they probably are not included in the MAF since the MAF does not include new noncity-style addresses from the DSF. If the new units in these smaller counties have city-style addresses then it is possible that the ACS does not use those new DSF addresses because of the increased chance that those DSF records duplicate existing records in the frame.

Two other areas of concern are the large counties (50,000 or more housing units) with a high growth rate where ACS counts are 3.8 percent higher than the housing units estimates for these counties and medium sized counties (10,000 to 50,000 housing units) with little growth (less than five percent) where ACS counts are 4.3 percent higher than the housing unit estimates (Devine et al., 2008).

Given that most of the growth in the ACS housing unit counts is from the DSF, obtaining county-level growth rate information may help the ACS more accurately determine where to use certain types of new records from the DSF.

Devine, et al. examined several alternative filters in an effort to determine if the MAF can be used to improve the county-level housing unit estimates process. While alternative filters did reduce the differences between the housing unit estimates and the ACS housing unit counts at the county-level, that does not necessarily mean that the ACS should adopt those filters. These alternative filters would exclude some categories of units where the majority of units have been shown to be valid. Two such examples are EDS records and Census deletes that persist on the DSF. Both of those categories of records will be discussed in later sections of this report.

The Devine and Reese reports did not specifically address coverage because there was no measure of "truth" to which comparisons could be made. However, the FACHS National Evaluation used block listings to determine ground truth and to produce coverage estimates. They estimated gross undercoverage in high growth blocks at 8.8 percent. The largest component of gross undercoverage was omissions (6.2 percent). They also measured gross overcoverage for high growth blocks at 13.9 percent. Since most of the growth reflected in the MAF is provided by the DSF, this suggests that we may be adding more than just new residential units in high growth areas.

Identifying areas of high growth and improving coverage in those areas should be a priority for the ACS.

3.1.4 Urban/Rural Areas

The research indicates that coverage in rural areas is worse than in the urban areas.

The FACHS National Evaluation showed that gross undercoverage is higher in the rural areas and that omissions account for 78.6 percent of that undercoverage (Li, 2008). While filter rules can be examined and potentially modified to account for erroneous exclusions, improving overall coverage in rural areas cannot be done without finding a way to get missing units added to the MAF.

Undercoverage for mobile homes is also a problem. The FACHS National evaluation estimated the gross undercoverage for mobile homes at 18.9 percent, including an omission rate of 15.2 percent. The undercoverage rate for mobile homes in the current surveys' area frame was higher than the undercoverage rate of those units in the permit frame (24 percent versus 15 percent), but coverage of mobile homes appears to be an issue regardless of the type of area where the mobile home is located.

Field operations are one way to identify missing units in the rural areas and add them to the MAF. Almost 94 percent of the adds from the DAAL listings from September 2003 through the fall of 2004 were not found on the DSF in 2004 (Perrone, 2005). This is a good indication that the DAAL operation is successful at finding units missing from the MAF and adding them.

Although undercoverage in rural areas is not surprising given that the DSF does not update noncity-style addresses, it is somewhat surprising that gross overcoverage is higher in rural areas than in urban areas (Li, 2008). The Area Frame Study and the FACHS National Evaluation identified duplication as a major source of overcoverage in the rural areas. Both studies found that duplication was higher in areas with some indication of E-911 address conversions. While not exclusive to rural areas, address conversions do occur more frequently in those areas. Overcoverage in rural areas is not limited to blocks with E-911 address conversions. Gross overcoverage in blocks with no E-911 activity was still almost ten percent (Li, 2008). In areas with some indication of E-911 address conversions, the overcoverage is most likely due to new city-style addresses on the DSF that duplicate existing noncity-style addresses already on the MAF.

Another potential source of overcoverage in the rural areas, particularly in blocks listed by DAAL, are units deleted by DAAL. Deletes accounted for 13.4 percent of all DAAL actions (Perrone, 2005). However, the ACS does not automatically remove those deleted units from the sampling frame because ACS has obtained interviews for some of those units. DAAL has a limited search area while the ACS can search for the unit anywhere in the county. The theory is that these units were outside the DAAL search area and may exist elsewhere in the county. There are, of course, some DAAL deletes that do not exist on the ground. Since we cannot distinguish the true nonexistent cases from the cases deleted due to geocoding error, the nonexistent units remain in the ACS sampling frame contributing to overcoverage.

While the 2010 Census operations should improve coverage for these areas, the coverage will degrade as we get further away from the Census unless we can find a way to pick up growth without increasing duplication.

3.1.5 New Construction

The FACHS National Evaluation estimated the gross undercoverage for new construction to be 21.7 percent with an omission rate of 18.9 percent. Undercoverage is even worse in area frame blocks where the undercoverage is near 50 percent and omission is about 40 percent.

The FACHS New Construction Study for 2008 found that about 94 percent of the current survey permit frame addresses interviewed in November 2005 for two surveys were later found on the January 2007 MAF extracts. This indicates that new construction is being picked up on the MAF but that there is a lag between the time construction is completed and the time that the new construction addresses appear in the MAF. This is not surprising given the processing steps that must occur between the time that the Census Bureau receives the DSF from the USPS and the time that the ACS receives those addresses for use in sampling.

The FACHS Permit-DSF Lag study measured the time between when a permit was issued and when the address appeared on the DSF. The most relevant finding for the ACS was that, on average, permit units appeared on the DSF at about the same time that construction was completed (Flanagan, 2007). In addition, approximately 25 percent of the permit frame addresses were found on the DSF before a permit was issued. The FACHS 2008 New Construction Study reported similar results.

Much of the new construction probably will not be reflected in the same sample year that construction is complete due to the lags associated with the various processing steps involved in getting new addresses from the DSF into the ACS sampling frame. For example, if a new unit is completed in January of the current year and appears on the DSF for the first time that same month, then that unit will not be included in the ACS sampling frame until July of that year. This is because the MAF is updated with the DSF once in the spring and once again in the fall. In fact, even if that unit appeared on the DSF the previous October–three months before construction was completed--it still would not appear on the sampling frame until July. The frame created in July, however, is used to select the sample for the following year. So although construction of the unit in this example was completed in January of the current year, and may have even appeared as early as the October DSF the previous year, processing lags result in the unit being excluded from the ACS until the year after construction was completed.

3.2 Research Question #2. Are there categories of MAF addresses whose quality has never been assessed?

3.2.1 Ungeocoded DSF Records

There is no documented research that examines the validity of the ungeocoded addresses from the DSF in the ACS housing unit frame or to determine if there is a better way to identify areas where the ACS should or should not use these ungeocoded DSF records. The ACS currently uses ungeocoded records in counties that contained any mailout/mailback areas in the 2000 Census, even if the majority of the county is considered to be in the duplication zone[3]. This leads to overcoverage in areas where those new ungeocoded DSF addresses duplicate existing records that do not have a city-style address and therefore cannot be matched or linked together. We have the opposite problem in counties that had no mailout/mailback areas for the 2000 Census. In those counties we exclude the ungeocoded DSF records. There is undercoverage in those counties where the new ungeocoded DSF addresses represent new growth.

The ACS typically excludes all ungeocoded new DSF addresses in those counties but has begun using ungeocoded new DSF records in certain counties where the risk of duplication is deemed to be low because of the high percentage of city-style addresses in those counties.

The USPS' Locatable Address Conversion System (LACS) file contains addresses that convert from noncity-style to city-style. While this allows for unduplicating units with a noncity-style address such as a Post Office Box or Rural Route/Box during the MAF updating process, it does not allow unduplicating units with physical descriptions.

Having a block geocode improves our ability to accurately filter new DSF addresses. Methods for imputing block codes for ungeocoded records could be investigated.

3.2.2 Noncity-Style Addresses on the DSF

DSF records without a city-style address are not added to the MAF so they do not really fall into the category of "MAF addresses." However, since these addresses are on the DSF they may be able to help us identify counties where the number of these addresses is increasing. The research has shown that there is substantial undercoverage of new construction addresses in rural areas and that there is undercoverage in small, fast-growing counties. The noncity-style addresses from the DSF could be used to identify these counties with the goal of improving coverage, perhaps by targeting those counties for additional field listings.

[3] The duplication zone is an area where DSF records are generally not used because of the chance those records duplicate existing records on the MAF. The duplication zone is discussed in more detail in section 3.3.2.

3.2.3 Puerto Rico

Coverage in Puerto Rico has not been investigated but there is undoubtedly significant undercoverage of new construction addresses in Puerto Rico given that the DSF addresses in Puerto Rico are not currently used to update the MAF.

3.2.4 Group Quarters

While outside the scope of this report, Group Quarters (GQ) coverage could be a concern. (Note that had there been any research on GQs to examine we may have included GQ coverage as part of this report.) The number of GQs in the ACS GQ frame continues to decline because there are very few sources of new GQs. Operations are deleting GQs but very few operations add GQs.

3.3 Research Question #3. Did previous research identify categories of MAF addresses that could benefit from further research?

3.3.1 Excluded From Delivery Statistics (EDS) Records

Several reports have investigated various subcategories of DSF addresses. One of the more scrutinized categories of DSF records includes records that the USPS excludes from their delivery statistics, or EDS records. These are records for addresses that are not current mail delivery points in the DSF. Some of these EDS records are likely to be incomplete new construction where mail is not yet being delivered. These are the ones we would like to include in the frame but we can not accurately identify them using available data from the DSF or MAF. We try to weed out the bad EDS records. Geography division (GEO) is able to identify, and ACS can exclude, the following categories of "bad" EDS records:

1. multi-unit placeholder records
2. ZIP code conversions
3. old records from LACS conversions
4. commercial/governmental units
5. potential college addresses

There are other EDS records that should be invalid for the ACS that can not be identified (e.g., postal service centers).

The 2007 National Estimates of Coverage Report determined that excluding EDS records from the ACS housing unit counts resulted in a 0.3 percent net undercoverage for the nation, as opposed to the 4.1 percent net overcoverage in the ACS frame with the EDS records. Devine performed a similar analysis comparing the housing unit estimates to ACS counts without EDS records and found that the difference between the ACS count and housing unit estimates decreased from 4.2 percent to 0.6 percent in 2006 when excluding these EDS records. Devine also found that excluding EDS records from the ACS frame had a larger impact on the high growth counties, providing further evidence that at least some of the EDS records represent new construction.

In the FACHS Filter Rules Research, Martin estimated that about 46.8 percent of EDS records are invalid compared to 15.4 percent of the DSF records that are Included in Delivery Statistics (IDS). IDS records are supposed to be valid current mail delivery points in the DSF. Excluding all EDS records from the ACS would bring the housing unit counts for the ACS closer to the housing unit estimates and other estimates of truth and would decrease the overcoverage in the ACS housing unit sampling frame; however, it would also increase the gross undercoverage in the ACS sampling frame, particularly for new construction units.

It is worth investigating whether or not there are subcategories of these EDS records that may be more likely to represent invalid units than others. Martin looked at the rate of invalid EDS records by delivery point type. While there were some delivery point types that were invalid more often than others, each classification of delivery point type did also contain a large percentage of valid units.

Colosi also calculated validity rates for EDS records by delivery point type using ACS outcome codes. While the results were similar for some delivery point types, they were very different for others. As Colosi suggested, this could be due to problems associated with using ACS outcome codes to determine validity status.

Although Field Representatives (FRs) can identify valid EDS records at time of interview, overcoverage on the ACS frame results in increased interviewing costs. Although a lot of research focused on the EDS records, these records could still use some additional scrutiny since they appear to be a major contributor to overcoverage on the ACS frame. More research should be done to determine if subcategories of EDS records could be eliminated from the frame to reduce overcoverage without increasing the already high undercoverage of new construction units.

3.3.2 New DSF Records in the Duplication Zone

"Duplication zone" is a phrase used by DSMD to describe areas where the new DSF records are more likely to duplicate existing records that do not have a city-style address. The duplication zone is defined using an address characteristic type (ACT) code that is assigned to each block in the country every year. The ACT code categorizes blocks according to the types of addresses in the block (city-style, noncity-style, mixed, etc) and whether or not there is any DSF coverage in the block. The duplication zone also includes entire counties that did not contain any mailout/mailback areas for the 2000 Census.

The FACHS Filter Rules Research examined rates of invalid new geocoded DSF records by ACT code and found that the current definition of the duplication zone used by the ACS is effective at minimizing the amount of duplication introduced into the frame by using new DSF addresses in these areas. The ACS currently uses new DSF records in blocks containing a combination of city-style and noncity-style addresses where the percentage of city-style addresses is 85 percent or more. There are also some blocks with 80 percent to 85 percent city-style addresses where the ACS uses the new DSF addresses. The FACHS Filter Rules Research provided some evidence suggesting that the duplication zone could be narrowed somewhat to allow the use of new DSF

records in blocks where 75 percent or more of the addresses are city-style. The standard errors of the estimates were high so more research is needed before deciding to expand the use of the DSF in these areas.

3.3.3 Census Deletes that Persist on the DSF

A third subcategory of DSF addresses examined in the FACHS Filter Rules Research are records that were deleted by 2000 Census operations but remain as residential records on the most recent DSF. This category of records is thought to include new construction that was incomplete as of Census Day 2000 as well as other records that were ineligible for inclusion in the 2000 Census but that now exist. Approximately 36 percent of these units were found to be invalid. Further examination into these census deletes that persist on the DSF may help determine if there is a subcategory of them that could be excluded so that we could reduce the overcoverage without causing an unacceptable increase in undercoverage.

3.3.4 Counties With High Growth Rates

POP research identified large counties with high growth rates as a category of addresses where ACS counts are much higher than the housing unit estimates. The ACS counts were much lower than the housing unit estimates in small counties with high growth. Assuming good growth rate data can be found, it could be worthwhile to investigate the use of growth rate data to identify areas where we should or should not use new DSF addresses.

3.4 Research Question #4. Did previous research identify categories of MAF addresses whose coverage should be improved?

All of the categories of addresses where coverage needs to be improved have been discussed in previous sections so this section will briefly summarize those findings.

3.4.1 Addresses in Rural Areas

There is a substantial amount of undercoverage in rural areas, particularly rural areas with growth. According to the FACHS National Evaluation, about half of the new construction addresses were missing from DSMD's area frame blocks A high percentage of this undercoverage is due to records missing from the MAF. There is also high overcoverage in rural areas, much of which could be duplicates added by the DSF.

3.4.2 Mobile Homes

The gross undercoverage rate for mobile homes was estimated at about 19 percent compared to less than six percent for conventional housing units (Li, 2008). The segment of the population residing in mobile homes is likely to be different than those living in conventional housing so this undercoverage will most likely result in biased estimates.

3.4.3 High Growth Areas

Identifying areas of high growth could be useful to the ACS since those areas have both high gross overcoverage and undercoverage. Identifying areas of high growth could also allow us to target those areas for future field work, particularly in high-growth rural areas where the omission rate is high.

4 Conclusions

There appears to be net overcoverage in the ACS housing unit frame. The ACS housing unit counts are consistently higher than the housing unit estimates and various measures of "truth." National gross overcoverage was estimated at 10.2 percent and gross undercoverage at 6.4 percent. The largest component of undercoverage is omissions, or units that are missing from the MAF altogether. Although some of the differences between the MAF the various benchmarks can be explained, some overcoverage and undercoverage probably still exists.

Undercoverage is high in rural areas, particularly in areas experiencing growth. Some of the growth in these areas is not being picked up and added to the MAF because the MAF updating process excludes new records from the DSF that do not have a city-style address. One report measured undercoverage of new construction in rural areas at nearly 50 percent, including an omission rate of about 40 percent. Without reducing omissions, coverage in these areas will continue to degrade.

An examination of the number of noncity-style addresses in the DSF that are excluded from the MAF could help us identify fast growing rural areas of the country. These areas could be targeted for field work to pick up new addresses.

Undercoverage of new construction exists outside of rural areas as well. There is evidence that the DSF does eventually include much of the new construction, but until the DSF catches up, there will be undercoverage of new construction in the ACS. Processing lags from the time a new construction address appears in the DSF until the time the ACS uses that address also result in undercoverage.

Mobile homes also have a high undercoverage rate, regardless of whether they are located in rural areas or other areas.

Overcoverage is likely due to the inclusion of certain categories of DSF records; however, it may be difficult to remove whole classes of DSF records without increasing the already high undercoverage rate of new construction addresses. The use of various classes of DSF addresses merits further research.

Further research into the definition of the "duplication zone" should also be done to help identify areas where the ACS should expand or discontinue the use of new DSF addresses. Incorporating county-level growth rates in the definition of the duplication zone should be studied.

Results from the address canvassing operation will be incorporated into the MAF extracts that GEO will deliver to the ACS in January 2010. These extracts could be used to assess various categories of addresses. Using the address canvassing results as ground truth could allow us to determine if there are types of records in the ACS housing unit frame that are more likely to represent overcoverage than others. At the same time we can examine the addresses added by address canvassing that matched to other records in the MAF to determine if there are categories of addresses in the MAF being excluded by the filter that we should be including.

5 Limitations

One limitation that is consistent throughout the various research reports is the lack of a "true" count of housing units at the various levels of geography that could be used to measure coverage. The 2007 National Estimates of Coverage report applied the rate of change in the housing estimates to the dual system estimates of housing unit coverage from the Census 2000 Housing Unit Coverage Study to determine truth. The FACHS National Evaluation used block listings to determine the true housing count in those blocks. Research by POP did not attempt to determine truth but rather sought to explain the differences.

Another limitation is that several reports use different reference dates. The ACS filter has been revised several times since some of the research was undertaken so the findings may be slightly dated.

However, the goal of this report is to determine what the research means in terms of coverage for the ACS and not to compare the various reports or to calculate actual coverage rates for the ACS. The results of the various research reports found similar trends in coverage, so any inconsistencies in definitions should have minimal impact on the findings in this report.

6 Contact

Please contact Larry Bates (301-763-5926) or Jim Hartman (301-763-1976) of the Community Address Updating Systems Branch in the Decennial Statistical Studies Division with questions about this report. You may also contact them via email at Lawrence.Martin.Bates@census.gov or James.E.Hartman@census.gov, respectively

7 References

Colosi, R., Kwiat, A. and Perrone, S., 2006. "Evaluation of Two Subcategories of the Delivery Sequence File," *2006 Proceedings of the American Statistical Association,* Government Statistics Section [CD-ROM], Alexandria, VA: American Statistical Association

Corlett, M., Kennel, T., Flanagan, T., Liu, X. and Loudermilk, C., 2006. *Report from Frame Assessment for Current Household Surveys – Area Frame Study,* Demographic Statistical Methods Division 2010 Sample Redesign Document # 2010-1.0-G-3

Devine, J., Reese, A., Robinson, J.G. and West, K., 2008. Investigating the Use of Data from the Master Address File for Developing Housing Unit Estimates

Flanagan, T., Loudermilk, C. and Liu, X., 2007. *Analytical Report – Frame Assessment for Current Household Surveys – New Construction Study Supplement (FACHS-NCS-SUPP) (Permit-DSF Lag Study),* Demographic Statistical Methods Division 2010 Sample Redesign Document # 2010-4.0-G-4

Flanagan, T., Loudermilk, C. and Liu, X., 2008. *Frame Assessment for Current Household Surveys: New Construction Study for 2008 (FACHS-NCS-2008),* Demographic Statistical Methods Division 2010 Sample Redesign Document # 2010-4.0-G-2

Johnson, N., Hakanson, A., Tomaszewski, C. and Clark, S., 2008. *Address Coverage Improvement and Evaluation Program – 2007 National Estimate of Coverage of the Master Address File*

Li, M., Loudermilk, C. and Liu, X., 2008. *Frame Assessment for Current Household Surveys (FACHS) National Evaluation: Final Analytical Report*, Demographic Statistical Methods Division 2010 Sample Redesign Document # 2010-4.0-G-12

Martin, J. and Loudermilk, C., 2008. *Frame Assessment for Current Household Surveys (FACHS) Fdilter Rules Research: 2008 Analytical Report*, Demographic Statistical Methods Division 2010 Sample Redesign Document # 2010-4.0-G-8

Perrone, S., 2005. *Address Coverage Improvement and Evaluation Program – Assessment of the Demographic Area Address Listing Program*

Reese, A., 2006. *A Comparison of Housing Unit Estimates to the American Community Survey Aggregate Master Address File,* Presentation and Paper for the Southern Demographic Association Annual Conference, Durham, NC (November)