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

February 27, 2013

2010 CENSUS PLANNING MEMORANDA SERIES

No. 253

MEMORANDUM FOR The Distribution List

From: Burton Reist *[signed]*
Acting Chief, Decennial Management Division

Subject: 2010 Census Evaluation of Address Listing Maintenance Using Supplemental Data Sources

Attached is the 2010 Census Evaluation of Address Listing Maintenance Using Supplemental Data Sources. The Quality Process for the 2010 Census Evaluations, Experiments, and Assessments was applied to the methodology development, specifications, software development, analysis, and documentation of the analysis and results, as necessary.

If you have questions about this report, please contact Christine Gibson Tomaszewski at (301) 763-1997 or Kevin Shaw at (301) 763-1851.

Attachment

2010 Census Evaluation of Address List Maintenance Using Supplemental Data Sources

U.S. Census Bureau standards and quality process procedures were applied throughout the creation of this report.

Christine Gibson Tomaszewski
Decennial Statistical Studies Division

This page intentionally left blank.

Table of Contents

List of Tables

List of Figures

Executive Summary

This evaluation examines whether using information available to the U.S. Census Bureau, but not currently incorporated into the address list updating process, can improve the address list for future censuses and surveys. The Decennial Statistical Studies Division evaluated four topics that presented potential for improving address list maintenance activities:

Topic 1. Geocode Imputations
Topic 2. Administrative Records
Topic 3. Local Geographic Information Systems Coordinates
Topic 4. American Community Survey Time of Interview Actions

The methodology to evaluate each topic included the 2010 Census Address Canvassing results and the final census status. Data from a supplemental field operation assisted in evaluating the geocode imputation topic. The address list refers to the Master Address File/Topologically Integrated Geographic Encoding and Referencing database. Extracts from the address list were used to obtain the 2010 Census Address Canvassing results. The final census status came from the Census Unedited File.

Topic 1. Geocode imputations: How can we use geocode imputations to improve address list maintenance? How accurate are imputed geocodes for ungeocoded address list records?

When planning this research, the Geography Division had licensed a software package (GeoStan) that imputed latitude and longitude coordinates and geocodes, or census blocks, for addresses. This evaluation demonstrated a strong link between the geocode imputation type and the accuracy of geocodes derived from latitude and longitude coordinates. The analyses conducted found that imputed latitude and longitude coordinates for an address often existed outside of the corresponding imputed geocode, so it was necessary to re-create the imputed geocode based on the Geography Division's official, public Census 2000 tabulation shape files.

For this evaluation, a study universe of 6,440,043 ungeocoded address list records was identified. The 2010 Census Address Canvassing operation found 2,844,428 address records with imputed geocodes. By imputation type, the results are:

- 97 percent (1,033,090/1,064,701) of Point geocode imputations, based on latitude and longitude coordinates associated with structures or land parcels, were found to be accurate;
- 81 percent (760,164/937,615) of Best Address geocode imputations, based on street names and house number rangers, were found to be accurate;
- 56 percent (112,228/247,790) of Other Address and Best Zone Improvement Program (ZIP) Code geocode imputations, based on street names and house number ranges or the center of a nine-digit ZIP Code area, were found to be accurate;
- 12 percent (72,328/594,322) of other ZIP Code geocode imputations, based on the center of a seven-digit, five-digit, or three-digit ZIP Code area, were found to be accurate.

Further, to remove the bias associated with the above overlapping 2010 Census Address Canvassing operation universe, a supplemental field operation was designed and conducted using

Locate Address Assignments. The Locate Address Assignment method of locating (searching for) a particular address in a county was designed into the Automated Listing and Mapping Instrument software. This evaluation was the first usage of Locate Address Assignments in an automated instrument; thus form design, printing, and keying were not necessary for this evaluation. The supplemental field operation was split into two phases to allow staff more time to learn about the Locate Address Assignments and change the procedures if necessary. The first phase of the supplemental field operation consisted of 100 address records per Regional Office, concentrated in two to five counties to allow easier coordination of this new type of work assignment. After a debriefing, the second phase of the supplemental field operation searched for the remaining sampled address records.

The results of this supplemental field operation confirmed the reliability of the Point and Best Address geocode imputations. The nationally representative, field-validated sample of 6,549 address records showed, at the 90 percent confidence level:

- 90.6 percent (89.2, 92.1) of Point geocode imputations were accurate, and
- 73.4 percent (71.9, 74.8) of Best Address geocode imputations were accurate.

Point and Best Address geocode imputations represent 62.1 percent of the study universe (3,998,531 of the 6,440,043 previously ungeocoded, and potentially valid, address records nine years after the last census) and would have represented a 2.8 percent (3,998,531/143,356,106) increase in the eligible 2010 Census Address Canvassing workload if the imputed geocodes had been applied to the address list prior to creating the dependent list for the operation.

Topic 2. Administrative Records: How can we use Administrative Records to improve address list maintenance? Do administrative records provide intelligence for adding new records, confirming existing address list records, and/or identifying deletes or vacant units?

Administrative records do provide intelligence confirming existing address list records and identifying deleted and vacant units. For this topic, address list records were required to have a non-blank address type and a ZIP Code of length five for matching purposes. Due to these requirements, the distribution of 2010 Census Address Canvassing actions for this research topic is different from the total U.S. counts reflected in the 2010 Census Address Canvassing Assessment. The administrative records used in this topic were the 2007 and 2008 Statistical Administrative Records System, the most recent vintages available before the 2010 Census Address Canvassing operation.

Of the about 9.9 million adds in the 2010 Census Address Canvassing operation that were eligible for analysis in this evaluation, the 2007 and 2008 Statistical Administrative Records System contained 67 percent (2,741,141/4,071,530) of the adds that matched back to the address list and 5 percent (265,163/5,836,247) of the new adds. This confirms that administrative records data can provide independent evidence of address validity for the universe of address records excluded from the 2010 Census Address Canvassing filter. This administrative records data potential, if valid administrative records addresses could be identified in an efficient and accurate manner prior to the listing operation, would have represented a possible addition of over 2 percent (3,006,304/143,356,106) to the eligible 2010 Census Address Canvassing operation workload.

Identification of potential adds that do not meet the current filter rules or are not in the address list could focus efforts to improve the address list on records, or areas containing records, that might be targeted for field operations or other address list maintenance activities. For instance, if these potentially valid records are ungeocoded, they could be processed through imputation software (see previous research topic) to see if sufficiently high quality imputed geocodes are available. Alternatively, the Census Bureau could request governments with significant numbers of these records to consider submitting address files with local Geographic Information Systems coordinates (see next research topic).

In order to assess whether or not administrative records data provide intelligence for confirming existing address list records, records with positive 2010 Census Address Canvassing operation action codes were matched to administrative records. The 2007 and 2008 Statistical Administrative Records System contained: 88 percent (85,909,962/97,511,328) of all 2010 Census Address Canvassing verify address records; 88 percent (4,813,958/5,445,667) of all 2010 Census Address Canvassing move address records; and 69 percent (12,639,561/18,444,809) of all 2010 Census Address Canvassing change address records. This confirms that administrative records data can provide independent evidence of address validity for the universe of address records included from the 2010 Census Address Canvassing filter. The 2010 Census Address Canvassing operation cost nearly 845 million dollars including field and contract overhead costs (Holland, 2012), and therefore offers high cost reduction potential. This administrative records data potential, if valid administrative records addresses could be identified in an efficient and accurate manner prior to the listing operation, would have represented a possible reduction of over 72 percent (103,363,481/143,356,106) to the eligible 2010 Census Address Canvassing operation workload.

To assess whether or not administrative records data provide intelligence for identifying addresses that should be removed from the address list, records with negative 2010 Census Address Canvassing operation action codes were matched to administrative records. Of all address records that received negative action codes in the 2010 Census Address Canvassing (delete, duplicate, nonresidential, uninhabitable), 73 percent (14,521,685/19,889,845) were not present in the 2007 and 2008 Statistical Administrative Records System. In particular, 74 percent (10,589,737/14,692,445) of address records with 2010 Address Canvassing delete actions were absent from both the 2007 and 2008 Statistical Administrative Records System. This administrative records data potential, if invalid address list addresses could be identified via administrative records in an efficient and accurate manner prior to the listing operation, would have represented a possible reduction of over 10 percent (14,521,685/143,356,106) to the eligible 2010 Census Address Canvassing operation workload.

The 2010 Census Address Canvassing filter offered an 83.5 percent (153,866,943/184,300,099) correct prediction of the final 2010 Census status, while the 2007 and 2008 Statistical Administrative Record System match status offered an 80.7 percent (148,705,784/184,300,099) correct prediction of the final 2010 Census status. This small net difference of 2.8 percent in predicting the correct final 2010 Census status indicates a strong similarity in predicting the final 2010 Census status between the Statistical Administrative Record System match status and 2010 Census Address Canvassing filter. The results showed that for 65 percent (119,427,271/184,300,099) of the address list, and for 79 percent (119,427,271/151,199,426) of

the 2010 Census Address Canvassing workload, a corresponding Statistical Administrative Record System address existed.

The 2010 Census Address Canvassing filter included Local Update of Census Addresses records in order to offer feedback to participating governments, but resources may not permit such a large field operation for future decennial censuses, so an alternative method of feedback may be needed. Since the Statistical Administrative Record System match status and the 2010 Census Address Canvassing filter had similar outcomes through the 2010 Census, the Statistical Administrative Record System match status offers potential to support alternative methodologies to a national field operation.

In summary, the results in this evaluation show that the presence of addresses in the Statistical Administrative Record System increases the likelihood for positive listing actions, and the absence of addresses in Statistical Administrative Record System increases the likelihood for negative listing actions.

Topic 3. Local Geographic Information Systems coordinates: How can we use local Geographic Information Systems coordinates to improve address list maintenance? Were the geocodes derived from local Geographic Information Systems coordinates accurate?

Between 2007 and 2009, the Geography Division used local Geographic Information Systems coordinates to identify geocodes for ungeocoded addresses (geocode adds, or new geocodes) and to verify or correct geocodes on existing geocoded addresses (geocode changes, or corrected geocodes) on the address list. This quantity of address records represents 2.6 percent (3,707,693/143,356,106) of the eligible 2010 Census Address Canvassing workload. The address list records with local Geographic Information Systems coordinate actions were validated in the 2010 Census Address Canvassing operation at a higher rate than other geocoded addresses, 75 percent (2,776,707/3,707,693) as compared with 68 percent (97,524,708/143,356,106). About 85 percent (3,139,944/3,707,693) of the records with local Geographic Information Systems coordinate actions were valid in the 2010 Census.

The vast majority of these locally updated address records had corrected geocodes, of which 90 percent were valid in the 2010 Census. The majority of the addresses with corrected geocodes sent to the 2010 Census Address Canvassing operation had a positive outcome, with 87 percent receiving verify or change actions. These 2.9 million addresses would have been moves during the 2010 Census Address Canvassing operation without their corrected geocodes. A move is a post-processing outcome requiring both a 2010 Census Address Canvassing add and a 2010 Census Address Canvassing delete action in the field operation. Addresses with corrected geocodes were more frequently validated (81 percent, or 2,702,442/3,358,784) than addresses with new geocodes (21 percent, or 74,265/348,909) in the 2010 Address Canvassing operation.

Half of the addresses with new local Geographic Information Systems geocodes (173,377/348,909) were deleted in the 2010 Census Address Canvassing operation, a higher delete rate than for all addresses eligible for the 2010 Census Address Canvassing operation at 10 percent (or 15,529,724/ 143,356,106). Ungeocoded records are often new construction, and addresses are often added to the Delivery Sequence File before their corresponding structures are built. Similarly, local governments often keep track of new residential addresses at the permit

stage (before construction begins). As of the 2010 Census Address Canvassing operation, 85 percent of the addresses with new geocodes were not yet on the Delivery Sequence File, which may help explain the high delete rate for addresses with new geocodes. Developing quality checks on addresses with new geocodes may be advisable given the high deletion rates found in the 2010 Address Canvassing operation.

The 2010 Census Unedited File gave results similar to the 2010 Census Address Canvassing action codes. Ninety percent (3,036,794/3,358,784) of the addresses with corrected geocodes were in the final 2010 Census universe, while only thirty percent (103,150/348,909) of the addresses with new geocodes were included in the final 2010 Census universe. Continued collection of local address files with associated Geographic Information Systems coordinates may lead to a smaller number of mis-geocoded units in the address list, which translates into a higher quality address list and reduced field costs due to mis-geocoded addresses.

Topic 4. American Community Survey Time of Interview actions: How can we use American Community Survey Time of Interview data to improve address list maintenance? How frequently were records with American Community Survey Time of Interview delete or other actions confirmed in the 2010 Address Canvassing Operation?

The purpose of this research topic was to investigate whether American Community Survey Time of Interview actions should be included in address filters. The American Community Survey filter currently uses American Community Survey Time of Interview delete actions to remove addresses from the American Community Survey sample universe; however, the intelligence of American Community Survey Time of Interview delete actions was not incorporated into the 2010 Census Address Canvassing filter. As recorded on the edited January 2009 American Community Survey Master Address File extract: there were 390,867 records with American Community Survey Time of Interview change actions; 93,760 records with American Community Survey Time of Interview delete actions; and 20,749 with American Community Survey Time of Interview nonresidential actions; amounting to 0.27 percent, 0.07 percent, and 0.01 percent of the 2010 Census Address Canvassing eligible workload respectively.

The majority (63 percent, or 245,811/390,867) of the addresses with American Community Survey Time of Interview change actions received positive Address Canvassing actions (verify, change, move, matched to add), and a similar percentage (62 percent, or 242,197/390,867) of records were in the final 2010 Census universe.

Over half of the addresses with American Community Survey Time of Interview deletes (56 percent, or 52,509/93,760) were also deleted during the 2010 Census Address Canvassing operation, for a total of 72 percent (67,864/93,760) that received negative 2010 Census Address Canvassing actions (delete, duplicate, nonresidential, uninhabitable) or no 2010 Census Address Canvassing action. Similarly, 77 percent, or 72,021/93,760, were not in the final 2010 Census universe.

The addresses with American Community Survey Time of Interview nonresidential action results were similar, overall, to the addresses with American Community Survey Time of Interview deletes, with 70 percent, or 14,453/20,749, negative 2010 Census Address Canvassing actions

(delete, duplicate, nonresidential, uninhabitable) or no 2010 Census Address Canvassing action, and 76 percent, or 15,856/20,749, that were not in the final 2010 Census universe.

The American Community Survey Time of Interview actions may be useful for address list maintenance, and should be considered in future targeting or data mining methods. While not an overwhelming association, the delete and nonresidential actions correlate with invalid records, while the change actions are associated with valid records.

Recommendations

Based on these results and the collective experiences and observations over the lifecycle of this research, the Decennial Statistical Studies Division presents the following recommendations:

(1) Conduct a Cost Benefit Analysis for Geocoding Activities
We recommend the Geography Division conduct a cost benefit analysis to assess the return on investment for licensing, renewing, and maintaining trained support staff to operate third-party geocoding software. The software is a substantial investment each year it is licensed, and foregoing license renewal for years when the software is unnecessary may incur additional charges; however, these costs are less than previous decennial address listing/canvassing activities. The Geography Division should quantify the quantities of address records geocoded by each of the internal geocoding activities, as well as development and maintenance costs of each (assuming internal geocoding activities continue). This formal documentation should quantify the error rates of external and internal geocoding activities and the tradeoffs of not conducting any geocode imputations or accepting a measured level of error in geocoding activities. Measuring the size and growth of the ungeocoded universe biannually (with the delivery sequence file updates) is essential. This biannual cost benefit analysis is necessary to inform budgeting and planning activities for the next Census.

(2) Apply Point and Best Address Geocode Imputations to the Address List as Warranted by the Cost Benefit Analysis
If the previous cost benefit analysis warrants purchase or renewal of the software, we recommend processing all ungeocoded address list records through the third-party software to capture only the Point and Best Address imputations. These geocodes need to be recalculated from the imputed latitude and longitude coordinates using Geography Division shape files. These final imputed, recalculated geocodes are reliable and accurate for address list updates. Both the geocode imputation and geocode imputation type (accuracy level) should be housed in the address list. We recommend that address list stakeholders be provided with documentation on the third-party geocode imputation software, and consider the usefulness of Other Address imputations and the Best ZIP Code imputations in specific applications. At this time, given the low level of accuracy, we do *not* recommend the use of Other ZIP Code imputations for geocoding ungeocoded records. This process should be repeated, as warranted by the biannual cost benefit analysis, for all ungeocoded records in the address list at that time. This universe will be largely comprised of new delivery sequence file adds and newly included in delivery statistics records.

(3) Periodically Evaluate the Accuracy of the Address List Geocode Imputations
Since address list maintenance procedures will change over time, and future address list input data sources will vary in both quantity and quality, we recommend periodic evaluation of any geocode imputations applied to the address list at intervals not to exceed two years. This evaluation work should be consistent with the analyses conducted in this evaluation, examining both the level of concordance with previous field activities and operations, and the level of accuracy measured by a supplemental field operation that investigates the universe of records for which the Census Bureau does not currently possess a geocode.

(4) Include Address List Geocode Imputations in Targeted Address Canvassing Research
The geocode imputations research results should be incorporated into the targeted address canvassing research. Geographical areas with highly accurate geocode imputations can be updated using imputation software rather than fieldwork. However, geographical areas with a majority of inaccurate imputations and numerous ungeocoded addresses are strong candidates for fieldwork targeted to locate the ungeocoded addresses and improve address coverage for those areas.

(5) Use Locate Address Assignments for Research and other Field Activities, Including Targeted Address Canvassing Research
The Locate Address Assignments worked well for our geocode imputation research, and are a useful method to collect data for field activities, research, and targeting focused on distinct addresses or groups of addresses, rather than a complete block listing. If only a few addresses in a block are in need of updating or of interest for research, Locate Address Assignments are a more efficient way of collecting information than listing every address in a block. The new corporate listing device should accommodate this type of assignment capability and utilize navigational routing.

(6) Construct A Single Integrated Administrative Record Data Repository
The process of acquiring, building, integrating, and conducting analysis on the large data files for the administrative records research topic required a substantial effort. If the matching and integration activities are done once at an agency level, similar to the integration and maintenance of the address list and the Business Register, all research and production activities would benefit significantly. Further, this integrated administrative records data repository should include regular address list updates so that users do not need to allocate significant amounts of time to merging this integrated administrative records data repository with the latest address list extracts. Identifying a full-time team of statisticians and IT Specialists to create and maintain a single integrated administrative record data repository which combines administrative records data with recent and historical address list extracts and decennial data will streamline this process.

(7) Expand Access to Administrative Records

All administrative records research and data acquired and integrated per the last recommendation should be made as widely available to the rest of the Census Bureau as possible. The results in this evaluation show that the presence of addresses in the Statistical Administrative Record System increases the likelihood for positive listing actions, and the absence of addresses in the Statistical Administrative Record System increases the likelihood for negative listing actions. Given these strong relationships, the Decennial Statistical Studies Division recommends providing Statistical Administrative Record System and other administrative records data accessibility to all areas in the Census Bureau conducting decennial census and current survey address list coverage and improvement research. All areas in the Census Bureau should have access to the Statistical Administrative Record System and other administrative records data for data modeling, data mining, and microsimulation research. Divisions involved in decennial census or survey operations should employ Title 26 trained personnel who are both authorized for and involved in active administrative records research with direct access to administrative records data in Title 26 secured areas. Having trained personnel with direct experience with the interactions between administrative records data and the decennial census or survey work of their division will optimize the development of specific methods to use administrative records data in census and survey operations. We recommend the pursuit of inter- and intra-agency approval for the following expansion of decennial census research using the Statistical Administrative Record System and other administrative records data: address frame filter research, targeted address canvassing research, improving Listing Quality Control methods, and general decennial census cost reduction research (for Address Canvassing and Nonresponse Followup).

(8) Update the MTdb with Administrative Records Flags

As measured in this evaluation, data from the Statistical Administrative Record System enhance the Census Bureau's ability to identify valid addresses, and are therefore useful for address frame updating and maintenance. Flags indicating whether an administrative records match existed for each address list record should be included on the address list, with annual updates of matches to the Statistical Administrative Record System and other administrative records data, similar to the existing Delivery Sequence File flags. Documentation detailing the meaning of each administrative records flag should be widely available for collaborative, cross-divisional research efforts by technical and subject experts to speed the development and implementation of practical administrative records use. These flags would be an ongoing method to incorporate administrative records data into decennial census and current survey work. The presence of a series of Statistical Administrative Record System match flags would allow address list users to include the presence or absence of these matches in their filter for specific applications, such as data modeling, data mining, and microsimulation work.

(9) Collect Local Geographic Information Systems Coordinates
As measured in this evaluation, quality, local Geographic Information Systems geocode corrections were highly reliable. Continued collection of these local Geographic Information Systems coordinates to verify or correct geocodes on existing geocoded addresses may lead to a smaller number of mis-geocoded units in the address list. This translates into a higher quality address list and reduced field costs due to mis-geocoded addresses. The costs of this continued collection program should be quantified for previous activities, and estimated for future ones to calculate the return on investment for this program.

(10) Collect Local Building Information along with Local Geographic Information Systems Coordinates
Half of the addresses with new geocodes derived from local Geographic Information Systems coordinates were deleted in the 2010 Census AC operation. A 'Provisional Add' unit status for an address with local coordinates that was not on the address list would allow storage of the new address's geocode, while filtering it out of surveys and the decennial census, so it does not increase field costs. Some local governments offer building footprint files online, so collecting information from local governments indicating whether an address record reflects an existing building or a building under construction (and a year expected to finish construction if known) may be possible and whether this information offers a good return on investment should be investigated. Alternately, existence on the Delivery Sequence File, alongside a delivery point type of 'Residential' and 'Included in Delivery Statistics,' could be used as a filter indicator for building existence. In an ongoing collection of addresses with local Geographic Information Systems coordinates, if local governments would include updated information on whether a building exists for a particular address record in a timely manner, then the address list would better reflect actual ground conditions.

(11) Further Investigate American Community Survey Time of Interview Actions
The American Community Survey Time of Interview actions are a small portion of the address list; and the results here are not decisive enough to require their inclusion in address filters at this time. Further research should be considered to improve understanding of these action codes, their characteristics, and their associated validity.

This page intentionally left blank.

1. Introduction

1.1. Scope

This evaluation will determine whether the United States (U.S.) Census Bureau can improve its address list for future censuses and surveys by using information available internally, but not currently incorporated into the updating process for the Master Address File (MAF). In the Census Bureau's Introduction to the 2020 Census (Bureau of the Census, 2009b), one of the potential design innovations was "supplementing the address frame with Federal or commercial administrative records." In the Census Bureau's Strategic Plan for the 2020 Census (Bureau of the Census, 2010), the strategies for "Goal 3: A Efficient 2020 Census" included the following items:

- "Use administrative records where feasible and publicly supported."
- "Leverage data, systems, and methods from the American Community Survey (ACS) and other surveys for incorporation into the 2020 Census design."
- "Investigate the feasibility of using additional high-quality sources of data or supplementing Census Bureau infrastructure with third-party methods and sources (such as USPS[1], ISP[2], SSA[3])."
- "Develop and validate performance and quality standards for third-party infrastructure and data sets."
- "Research and assess the adequacy of third-party infrastructure and data sources."

This evaluation used the results of the 2010 Address Canvassing (AC) operation and supplemental field operations to evaluate multiple research topics that showed potential for updating the MAF/Topologically Integrated Geographic Encoding and Referencing (TIGER[4]) database (MTdb).

Specifically, the Decennial Statistical Studies Division (DSSD) examined four research topics that presented potential for improving address list maintenance activities:

Topic 1. Geocode imputations
Topic 2. Administrative Records (AR)
Topic 3. Local Geographic Information Systems (GIS) coordinates
Topic 4. ACS Time Of Interview (TOI) actions

---

[1] United States Postal Service

[2] Internet Service Providers

[3] Social Security Administration

[4] TIGER is the mapping/spatial portion of the database, which contains street and other feature details, as well as mapspots, or geographic coordinates, corresponding to the location of MAF addresses.

1.2. Intended Audience

This 2010 Census Program for Evaluations and Experiments (CPEX) report is intended for managers and staff involved in planning and implementing the 2020 Census.

2. Background

2.1. The MAF/TIGER Database

The MTdb was the primary source of addresses for the address frame for Census 2000 and the 2010 Census, as well as many inter-censal surveys since Census 2000. The 2010 Census AC operation, conducted in spring 2009, used a filtered subset of geocoded address records from the MTdb to conduct a dependent listing of addresses across most of the nation and Puerto Rico (only remote areas of Alaska and Maine were excluded). Since its nationwide implementation in 2005, the ACS has used a filtered subset of both geocoded and ungeocoded address records from the MTdb to select their sample. The Address Control File for the 1990 Census and the United States Postal Service's (USPS's) Delivery Sequence File (DSF) were used to build the MAF in the late 1990's. Block Canvassing and Address Listing were the primary development mechanisms during Census 2000.

Since Census 2000, the MTdb has been augmented by a series of coverage improvement operations and address reviews by local and tribal governments; however, the main source for updating the MTdb has been the DSF. The DSF contains all addresses to which the USPS delivers mail, and has been the primary AR data source providing regular updates to the MTdb. The Geography Division (GEO) uses the DSF to update the MTdb twice per year. The GEO updates the MTdb with city-style[5] addresses from the DSF, but does not use non-city-style[6] DSF addresses due the risk of duplication resulting from not being able to match the addresses. In cases where both the non-city-style address and the city-style address can be found and a linkage provided in the Locatable Address Conversion System (LACS) File, GEO could link the old non-city-style address to the new city-style DSF address.

Aside from the biannual DSF updates to the MTdb, two ongoing operations regularly updated small numbers of addresses on the MTdb: Demographic Area Address Listing (DAAL) and the ACS. DAAL selects and lists block for current survey purposes and for the Community Address Updating System. According to the advance Office of Management and Budget (OMB) letter concerning DAAL 2012, 54,200 DAAL blocks were expected to be listed in fiscal year 2012, providing full funding was available, which is less than one percent of blocks nationwide.

---

[5] Addresses with both a house number and a street name are classified as city-style addresses.

[6] Addresses that do not include a house number and/or a street name are classified as non-city-style addresses. These can include General Delivery, Rural Route and Box number, Highway Contract Route and Box number, and Post Office Box addresses. They also include Location Descriptions, Targeted Non-Sheltered Outdoor Locations, Transient Locations, and crews of Maritime Vessels.

According to the ACS Design and Methodology (Bureau of the Census, 2009a), which discussed the 2007 ACS, "among the ACS sample addresses eligible for interviewing in the United States, … 41 percent were represented by CAPI[7] interviews. … Mailable addresses with neither a response to the mailout nor a telephone interview are sampled at a rate of 1 in 2, 2 in 5, or 1 in 3 based on the expected rate of completed interviews at the tract level. Unmailable addresses are sampled at a rate of 2 in 3. … Approximately 80 percent of CAPI cases require an FR[8] visit." Only these FR visits generate ACS TOI actions. The MTdb recorded ACS TOI change, nonresidential, and delete actions, but did not include dates for these actions at the time of this evaluation. Including ACS TOI actions from as far back as Fall 2004, 505,376 ACS TOI actions were recorded on the MTdb through January 2009.

To improve the accuracy of the MTdb in preparation for the 2010 Census, the GEO implemented the MAF/TIGER Accuracy Improvement Project (MTAIP). MTAIP focused on the spatial realignment of linear features in TIGER, such as streets and legal boundaries. This, in turn, should have improved the indication of the location of living quarters (LQs) in relation to the street features with the MTdb. There are a number of benefits derived from greater spatial accuracy of linear features in TIGER. First, geographic coordinates collected by field staff to represent the location of LQs on Census Bureau maps could more reliably assigning LQs to the correct census block for tabulation purposes. In addition, accurate location information facilitates navigating back to the LQ as needed, such as for interviewing for a survey or a decennial census.

Address files with associated GIS coordinate data accompanied the street centerline files submitted for the MTAIP by local governments. The GEO set up a program, the Structure Coordinate Enhancement of the MAF Addresses (SCEMA), to evaluate these address files and decided to process the local address files with associated GIS coordinates that passed their quality check. The SCEMA program processed these address files with local GIS coordinates to add addresses to the MTdb, identify geocodes for ungeocoded addresses, and to verify or correct geocodes on existing, geocoded addresses in the MTdb (Dixon and Tomaszewski, 2007).

2.2. Geocoding

Geocoding is the method where the Census Bureau assigns each MTdb address a numeric code (commonly referred to as a geocode or block code) identifying a geographic location. A geocode identifies the particular state, county, census tract, and census block where the address is located. The most common methodology for block-level geocoding of addresses provided by supplemental sources (e.g, the DSF) involves the use of address ranges (i.e., ranges of house numbers, such as 101-199, 201-299, etc.) assigned to sides of street segments in TIGER. If an address is built on a street that is not in TIGER or a street that does not have an associated house number range or the street has conflicting house number ranges, the Census Bureau is not able to associate the record to the census block associated with the right or left side of the street segment

[7] Computer-Assisted Personal Interviewing (CAPI)

[8] Field Representative

(i.e., geocode the address) in this manner. This is problematic, as the Census Bureau attempts to geocode city-style DSF addresses using address range geocoding.

Mis-geocoded, ungeocoded, or missing records in the MTdb may have resulted in higher costs for the 2010 Census AC and ACS operations. In particular, the 2010 Census AC dependent address lists excluded ungeocoded records. While this was due to operational constraints, it nonetheless prevented an opportunity to geocode a significant number of potential living quarters. The Census Bureau's National Estimate of Coverage reports indicated that the number of ungeocoded records on the MTdb have grown each year, between 2002 and 2009. According to Hakanson, Johnson, Tomaszewski, and Clark (2008),

> As we approach 2010, the universe of ungeocoded records continues to grow. One major result is that, for existing ungeocoded addresses on the Master Address File, the Address Canvassing operation in the 2010 Census will be expected to independently add these records. If Address Canvassing listers are not successful in finding and adding the ungeocoded units, many of the missed addresses will not be included in later census operations.

This also affected the ACS operations. According to Hakanson et alii (et al.) (2008), "In 2002, ungeocoded records amounted to 2.9 percent of the full ACS frame; this rose to 5.7 percent of the frame in 2007." The existence of an ungeocoded unit with a complete address in the DSF and MTdb does not mean the address exists on the ground. It was likely that some of these units were not added during the 2010 Census AC operation because they had yet to be built. In addition, some of these ungeocoded addresses may have represented current non-city-style MTdb addresses.

When this evaluation was proposed, the GEO had been studying a software package (GeoStan) that imputed latitude and longitude coordinates and geocodes for addresses. This software was a component of a larger software suite for various data cleansing and enhancement features. The GEO has not, however, used this software to update geocode information on the MTdb because the geocodes it provides have (prior to this report) not been fully evaluated. GeoStan derives flag settings indicating the level of accuracy for the geocode imputations. The level of accuracy is determined by how well the input address data matches or imputes the location using various reference files. GeoStan sets a flag to indicate a match at a lower level of geography (most precise) if the software can confidently match the input address information to a particular structure or successfully imputes the location using a street address range. Alternatively, GeoStan sets the level of accuracy flag to a higher geographic level (least precise) when a match to a particular structure or street address range imputation is not possible. For example, the Zone Improvement Program (ZIP) Code geocode imputations indicate a match accuracy at the ZIP code level because the software can only compute the centroid coordinates of a ZIP code.

In Holland (2009), the type of imputation for an address appeared to affect whether the address was found in the imputed geocode (census block) during the 2008 Census Dress Rehearsal (DR) AC operation. There were 10,346 geocode imputations for ungeocoded records available for the 2008 Census DR sites in North Carolina and California. The point imputations were most accurate during the 2008 Census DR AC with 99 percent, or 3,076/3,095, correctly imputed point geocodes. The best address imputations also exhibited a high level of accuracy during the

2008 Census DR AC with 89 percent, or 3,941/4,436, correctly imputed best address geocodes. Other address imputations (73 percent accurate or 22/30) and best ZIP Code imputations (73 percent accurate or 132/352) had similar levels of accuracy during the 2008 Census DR AC. The least accurate imputation types during the 2008 Census DR AC were the various other levels of ZIP Code centroids at 13 percent, or 288/2,298, correctly imputed other ZIP Code geocodes.

2.3. Assumptions

The key assumption for this study is:

- MTdb data extracts formed an adequate basis for estimating the number of living quarters nationwide at the completion of the 2010 Census AC operation for evaluating the following three research topics in this report: AR, Local GIS Coordinates, and ACS TOI Actions.

3. Methodology

3.1. Question to be Answered

This evaluation answers the following general research question:

How can we use additional information to improve address list maintenance?

This evaluation covered four research topics, and each was evaluated using a different methodology. The topics and their associated research questions are presented in order of the magnitude of their anticipated impact on address list maintenance. The research questions associated with each topic are:

Topic 1. Geocode Imputations
How can we use geocode imputations to improve address list maintenance? How accurate are imputed geocodes for ungeocoded MTdb records?

Topic 2. Administrative Records
How can we use AR to improve address list maintenance?
Do AR provide intelligence for adding new records, confirming existing MTdb records, and/or identifying deletes or vacant units?

Topic 3. Local Geographic Information Systems Coordinates
How can we use local GIS coordinates to improve address list maintenance? Were the geocodes derived from local GIS coordinates accurate?

Topic 4. American Community Survey Time of Interview Actions
How can we use ACS TOI data to improve address list maintenance?
How frequently were records with ACS TOI deletes or other actions confirmed in the 2010 Census AC operation?

3.2. Methods

The DSSD used the 2010 Census AC operation results to address the study questions in each research topic. The 2010 Census AC results were derived from the 2010 Census Initial Universe

Control & Management (UC&M)/Group Quarters Verification (GQV) extract[9], which was the first MTdb extract available after completion of the 2010 Census AC operation. The DSSD merged this post-AC file with other MTdb extracts to create the 2010 Census Address Frame Combination (Combo) file[10], which combined data from the 2010 Census AC operation extract[11], the post-AC file, and other extracts. DSSD used the 2010 Census Combo file to conduct the analyses of the AR, local GIS coordinates, and ACS TOI actions research topics. As a supplemental measure, the DSSD also used the 2010 Census Unedited File (CUF) to determine the final census status[12] of the records analyzed in each research topic.

Table 1 summarizes the methodology – each topic was evaluated using the 2010 Census AC results. A supplemental field operation assisted in evaluating Research Topic 1. DAAL listers carried out the national supplemental field operation using laptop computers via Locate Address Assignments (LAA) rather than the more common block listing assignments.

Table 1. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: Methods for Evaluating Address List Maintenance Research Topics

Research Topics	AC & CUF	Fieldwork
1. Geocode Imputations	✓	✓
2. Administrative Records	✓	×
3. Local GIS Coordinates	✓	×
4. American Community Survey Time of Interview Actions	✓	×

Sources: Pre-AC file, Post-AC file, 2010 Census Combo file, 2010 Census Unedited File, and GeoStan Imputations File.

The 2010 Census Combo file was not available for the initial Geocode Imputations analyses, nor the preparation for the supplemental field operation for that research topic, so the pre-AC file, the post-AC file, and the January 2009 ACS Supplemental File were used. For the same reason, the pre-AC file and the post-AC file were combined to prepare the MTdb input file for matching to the AR data. The output of the AR match was then merged back to the 2010 Census Combo file and the 2010 Census CUF for analysis.

The categorization of 'new adds' and 'adds matched to the MAF' was calculated differently from the 2010 Census AC Assessment. The magnitude of this difference was approximately 79 thousand addresses that were categorized as 'new adds' rather than 'adds matched to the MAF.' The 2010 Census AC Assessment considered new adds to be the records added during the regular 2010 Census AC operation that did not match to a records on the MAF or in the 2010

[9] This report refers to this file as the post-AC file. See Owens (2009b) for further details of the customer requirements for this MTdb extract.

[10] This report refers to this file as the 2010 Census Combo file. See Ward (2011) for further information of this database creation.

[11] This report refers to this file as the pre-AC file. See Owens (2009a) for further details of the customer requirements for this MTdb extract.

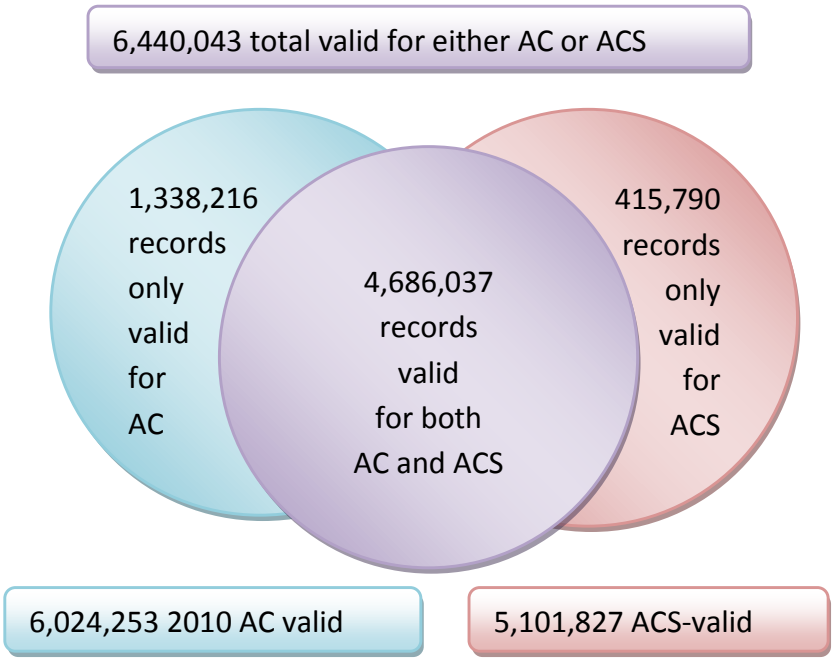
[12] The variable final_status on the CUF was used to determine the final 2010 Census status of each record.

Census Large Block AC (LBAC) operation (identified by the regular 2010 Census AC operation MAF source). The remaining adds were categorized as matches to existing records, including all adds originating in the 2010 Census (LBAC) operation (Address List Operations Implementation Team, 2012). For this evaluation, records that first came to the MTdb from either portion of the 2010 Census AC operation were considered new adds.

3.2.1. Geocode Imputations

DSSD used the 2010 Census AC results and a supplemental field operation to evaluate the accuracy of imputed geocodes. The results of this evaluation will inform the decision of if and how geocode imputations can be used to update the MTdb. If imputed geocodes are proven accurate and reliable, and these are applied to MTdb records, the quality of the MTdb will improve. Even though this quality improvement would come with some error, (some geocode imputations are inaccurate) for the MTdb user community this may offer a more accurate address list (with reduced undercoverage) from which more complete and accurate samples can be drawn.

Figure 1. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: Geocode Imputations Universe – 2010 Census AC Invalid Address Records due to a Missing Geocode and 2009 ACS Valid Ungeocoded Address Records



Sources: Pre-AC file, Post-AC file, January 2009 ACS MAFX, and GeoStan Imputations File.
Note that the above Venn Diagram is not to scale.

3.2.1.1 Universe Description

Address filters, or sets of business rules, identify valid records for the various ACS samples and identified potentially valid records for the 2010 Census AC operation. Using the edited January 2009 ACS MAF extract (MAFX)[13] and the 2010 Census AC extract[14], both the 6,024,253 otherwise-valid 2010 Census AC

[13] See Bates (2009) in the References for further details on the 2009 ACS filter.

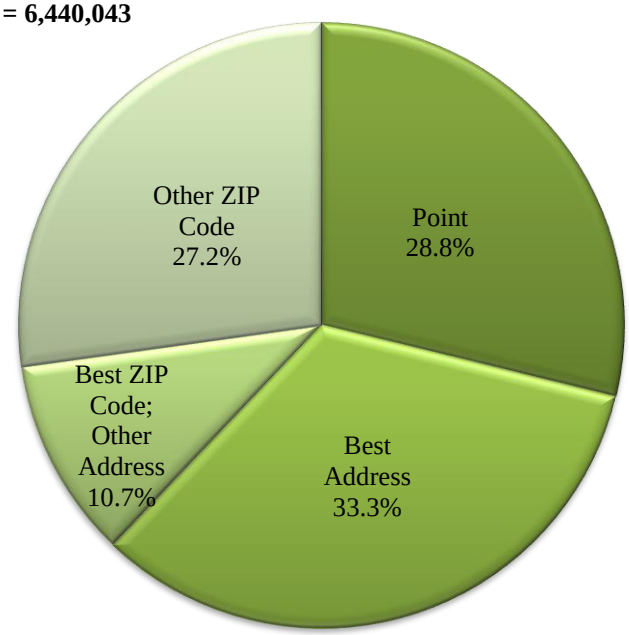
[14] See Owens (2009a) in the References for further details on the 2010 Census AC Operation filter.

ungeocoded records[15] and the 5,101,827 ACS-valid ungeocoded records comprised the universe of ungeocoded records for this research topic. These overlapping record sets yielded a combined universe of 6,440,043 addresses with valid geocode imputations. An imputed geocode was calculated for each of these records. As seen in Figure 1, the majority of the ungeocoded address research topic universe is valid for both address filters (2010 Census and 2009 ACS).

The distribution of the types of geocode imputations that comprise the ungeocoded records universe, shown in Figure 2, indicate that Point imputations represent a little more than one-quarter of the imputation types. The Point geocode imputations were derived from the latitude and longitude coordinates of the center of a parcel of land (usually designated/divided based on property ownership). Records were categorized as Point imputations when the internal GeoStan codes began with the characters 'AP.' Records with the internal GeoStan codes of 'AS0' or 'AS1'were categorized as Best Address imputations. Figure 2 shows that the Best Address imputations were the most common imputation type in the ungeocoded universe for this research topic at about one-third of the valid imputation types. The Best Address geocode imputations were derived by estimating the approximate location of a house number within the house number range (assuming equal intervals between each house number in the house number range) associated with an address's street. Records were categorized as Other Address imputations when the internal GeoStan codes began with an 'A,' but were not 'AP*,' 'AS0,' or 'AS1.' Records were categorized as Best ZIP Code imputations when their internal GeoStan codes began with 'ZB9,' indicating a ZIP Code centroid with block level accuracy. Other Address imputations and Best ZIP Code imputations (9-digit ZIP Code centroids with block level accuracy) are shown together in Figure 2, representing slightly more than one-tenth of the valid imputation types for this

Figure 2. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: Geocode Imputations Universe – Distribution of Geocode Imputation Types

N = 6,440,043



Sources: Pre-AC file, Post-AC file, January 2009 ACS MAFX, and GeoStan Imputations File.

[15] 'Otherwise-valid 2010 Census AC ungeocoded records' meant records that would have qualified for AC if they had been geocoded. Ungeocoded records did not go to AC, as AC was a block-based operation. Some examples of non-eligible records would have been non-city-style records or records with incomplete address information.

research topic. Records were categorized as Other ZIP imputations when the internal GeoStan codes began with 'Z' but did not begin with 'ZB9.' Other ZIP imputations contain various other levels of ZIP Code centroids. As shown in Figure 2, this imputation type is approximately 27 percent of the ungeocoded record universe for this research topic. These Other ZIP geocode imputations were derived from the latitude and longitude coordinates of the centroid (center) of a ZIP Code, which was often a very large geographic area containing numerous different census blocks. Since it is unlikely that stakeholders would approve including geocode imputations derived from other ZIP information in the MTdb, these types of imputations were not included in the sample chosen for further examination (via a supplemental field operation).

The software package that provides the geocode imputations offers both Census 2000 tabulation blocks and latitude and longitude coordinates. Initially, the DSSD examined the imputed Census 2000 tabulation blocks since the supplemental field operation required Census 2000 tabulation blocks and the 2010 Census AC results were available in Census 2000 tabulation blocks. However, the initial results for the Point imputations were poor. Examination of the data revealed that the imputed latitude and longitude coordinates for an address often existed outside of the corresponding imputed Census 2000 tabulation block. In other words, the software incorrectly assigned geocodes to imputed latitude and longitude coordinates. To correct this issue, the imputed latitude and longitude coordinates for each ungeocoded address were merged with the GEO's official, public Census 2000 tabulation block maps for each state in ArcMap[16] to create new imputed Census 2000 tabulation blocks. These ArcMap-revised imputed blocks were used for all analyses in the Geocode Imputations research topic.

Regional variation existed in the distribution of imputed geocodes. Texas had the largest amount (15.1 percent) of potentially valid ungeocoded records with imputed block codes. The District of Columbia had the smallest amount (less than 0.1 percent) of the total national potentially valid ungeocoded records.

The 2010 Census AC results supplied nationally representative accuracy statistics for the imputed geocodes. The 2010 Census AC results were examined by imputation type, census region, and state. The initial expectation was that accuracy in smaller geographies would vary from the national accuracy averages.

3.2.1.2 Sample Selection

The DSSD selected a unit-based, stratified sample from the combined universe of the 2009 ACS-valid and otherwise-valid 2010 Census AC ungeocoded city-style records for a supplemental field operation. This national sample was selected from the three most accurate types of geocode imputations, which are shown in Figure 2 and in Table 2. The universes eligible for sample selection and the subsequent supplemental field operation are given in Table 2. The sample was stratified by type of geocode imputation and whether records were found in the imputed block in

[16] ArcMap is a mapping software package that allows users to merge latitude and longitude coordinates with census blocks to provide geocodes. ArcMap is the name of the software and is not an abbreviation.

the 2010 Census AC operation, in a different block in the 2010 Census AC operation, or not found in the 2010 Census AC operation.

In designing the sample, the DSSD assumed a nine percent error rate for records found (or not found) in any block in the 2010 Census AC operation. The basis for this assumption stemmed from work that the Demographic Statistical Methods Division conducted, where two dependent block listings of a sample of tabulation blocks found a 16 percent inconsistency rate and a nine percent error rate (Kwiat, 2009). The 2010 Census AC Quality Profile reported a Dependent Quality Control check failure rate of over eight percent for assignment areas. While neither of these exactly corresponded to an error rate for 2010 Census collection blocks, the error rate of nine percent per block provided for a reasonable approximation.

Table 2. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: Sample Sizes[17] (with Universe Totals) by Strata – Geocode Imputation Type by 2010 Census AC Block Agreement

Geocode Imputation Type	Records Found in AC		Records Not Found in AC[18]	Total Records by Imputation Type
	Records Found in Imputed Block[19]	Records Found in Different Block[20]		
Point Imputations	158	437	564	1,159
	(1,008,461)	(30,175)	(818,010)	(1,856,646)
Best Address Imputations.....	158	1,334	1,076	2,568
	(732,760)	(173,284)	(1,235,841)	(2,141,885)
Other Address; Best ZIP Code	158	1,334	1,330	2,822
	(135,717)	(106,160)	(446,548)	(688,425)
Total by 2010 Census AC Block Agreement	474	3,105	2,970	6,549
	(1,876,938)	(309,619)	(2,500,399)	(4,686,956)

Note: Records not found in the 2010 Census AC operation will have lower accuracy than those found in the 2010 Census AC operation. Nine percent was used to approximate that difference.
Source: Sources: Pre-AC file, Post-AC file, January 2009 ACS MAFX, and GeoStan Imputations File.

The sample sizes shown in Table 2 were selected using the "probability proportionate to size" method with a 95 percent confidence interval. The first and third columns represent three percent precision, while the second column represents five percent precision. Since less than half of the imputed ungeocoded record universe was located during the 2010 Census AC

[17] Sample size $n = (t^2 * p * q) / (d^2)$ where t = 1.96 for a 95 percent confidence interval, d = 0.02 for accuracy of plus/minus 2 percent accuracy of our results, p is the probability of selection for a strata, and q is 1-p. Sample size includes a 25 percent oversampling factor.

[18] The probability for each stratum in the 'Records Not Found in AC' was the proportion of address records in each stratum that were accurately imputed during the national 2010 Census AC operation minus an estimated error rate of nine percent (discussed earlier).

[19] The probability for each stratum in the 'Records Found in Imputed Block' was 91 percent (100 minus the assumed nine percent error rate).

[20] The probability for the point imputations stratum in the 'Records Not Found in Imputed Block' was also 91 percent. The probability for the other two strata in the 'Records Not Found in Imputed Block' was 50 percent.

operation, difficulty in locating these addresses was anticipated. Thus, DSSD decided to increase the minimum sample size by 25 percent over the initial design parameters.

3.2.1.3 Supplemental Field Operation

The LAA method of locating (searching for) a particular address was designed into the Automated Listing and Mapping Instrument (ALMI) software early in the development of the ALMI, though it had been dropped from the training materials since LAA had not previously been used by FRs during DAAL. LAA allows FRs to search for a sampled address, with a county as the extent of their prospective search area. FRs can search a broader area with LAA than the assignment areas available during the 2010 Census AC operation. This evaluation was the first usage of LAA; thus form design, printing, and keying were not necessary for this evaluation. For this evaluative study, headquarters (HQ) staff examined a non-representative sample of cases in Prince George's County, Maryland in order to revise the training materials to prepare FRs for using the LAA method. This exercise determined that the LAA method could be used to conduct the evaluative work necessary for this research topic. Additionally, some training procedures were revised for addresses corresponding to an empty lot or in a gated community. Training materials were presented at the DAAL Supervisor's conference on June 13, 2011 in San Antonio, Texas. Groups of DAAL FR Supervisors, local FRs, and HQ staff used the training materials to find a non-representative selection of cases near the location of the DAAL Supervisor's conference on June 16, 2011. Debriefing sessions provided feedback on June 17, 2011.

After this LAA method testing in Maryland and in Texas, the production supplemental field operation was split into two phases to allow staff more time to learn about the LAA and change the procedures if necessary. The first phase of the supplemental field operation consisted of 100 records per Regional Office[21] (RO), and was fielded in July 2011. The records were concentrated in two to five counties per RO to allow easier coordination of this new type of work assignment. The counties with the largest number of sample records per county were selected for Phase I (to make up the initial 100 address workload, and if 100 in-sample records were not available in five counties, in-strata records were chosen to make up the difference). Phase I was originally planned for July 5, 2011 through July 15, 2011, however not all 1,200 cases were completed by the 15th of July. The supplemental field operation was extended through August 1, 2011. All cases were completed by July 27, 2011 and a debriefing was held on the final day with the RO supervisors to discuss the training materials.

[21] Regional Offices supervise and carry out fieldwork in their respective region of the country. At the time this fieldwork was conducted for this evaluation, there were 12 ROs. The RO workload distributions are available in the Appendix.

The actual start and finish dates of Phase I of the supplemental field operation are shown in Table 3.

Table 3. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: Geocode Imputations Supplemental Field Operation Schedule

	Start	Finish	Workload
DAAL Supervisors' Conference Testing in Texas	06/16/2011	06/16/2011	99
National Sample - Phase I	07/05/2011	07/27/2011	1,200
National Sample - Phase II	08/08/2011	09/30/2011	5,540

Source: Internal Correspondence Emails.

After Phase I, there were 5,540 address records remaining in the sample for Phase II. Phase II of the supplemental field operation was scheduled for August 15, 2011 through September 30, 2011. Since this was the end of fiscal year 2011, the end date could not be extended. In order to combine costs for other work done in Hawaii, the Los Angeles RO was given permission to start field work as early as August 1, 2011. However, due to difficulties in printing and shipping the revised training materials, the ROs started fieldwork between August 8th and August 23rd (calculated using the day of first completed cases for the date on which fieldwork started). Actual Phase II start dates for each RO are shown in the Appendix, alongside the Phase II workload and the overall workload for the national sample by RO. The overall Phase II start date, finish date, and workload are shown in Table 3.

The Appendix also contains the supplemental field operation workload and completed cases by RO. The RO with the largest sample was Dallas, which was consistent with Texas having the largest percentage of ungeocoded records with imputations in the nation.

The data files returned from the instrument containing the results of the supplemental field operation workload contained 6,432 records. Of the original sample, 117 address records were not completed (around two percent of the total national sample).

Analysis of the supplemental field operation used weighted numbers within strata sampling intervals as the base weights. The overall nonresponse rate was less than three percent, and the DSSD decided to adjust the responding population to reflect the sampled population. Nonresponse adjustment factors were calculated within each stratum based on the number of records sampled over the number of responders. The product of the nonresponse adjustment factor and the base weight provided the final weight.

The collected geocodes were compared to the imputed geocodes sent out for the supplemental field operation to determine whether geocodes had been correctly imputed. Incorrect imputations included both records that the supplemental field operation did not find and records that the supplemental field operation found in a different, non-imputed block. The supplemental field operation delete action codes were classified as incorrect imputations unless their unit status was demolished, in which case they were classified according to whether the returned geocode matched the geocode that was sent out. Only a few deleted records returned with a different, non-imputed geocode.

3.2.1.3 Logistic Regression

In order to investigate the net effect of geocode imputation type while accounting for other variables that significantly affect geocoding accuracy, we developed a logistic regression model. Census research and stakeholders suggest that variables such as census geography and multi-unit status might meaningfully affect geocoding accuracy (see Virgile, 2012). These categorical values were transformed into dichotomous variables for the logistic regression analysis in SAS. The data for this analysis are from the supplemental field operation sample. Non-response adjusted weights (representing the ungeocoded record count at the time of the sample) were used in this analysis. The dependent variable for the logistic regression was whether records were found in the imputed block (1) or not (0). The independent variables investigated were census region, multi-unit status, address filter validity, and imputation type.

Binary variables were created for three of the four census regions - Northeast, South, and West – that were used in the logistic regression analysis. The Midwest was the comparison or excluded category. If an address record was a multi-unit, it was assigned the value 1; otherwise, it was assigned the value 0. Thus, the comparison, or excluded, category consisted of single-unit address records.

As discussed in Section 3.2.2.1 Universe Description, the 2010 Census AC operation filter (Owens, 2009a) and the 2009 ACS filter (Bates, 2009) both provided records for this analysis. These two address filters used different rules to arrive at overlapping sets of address records. Agreement between the two filters might be an indicator of addresses that are more easily geocoded than addresses that failed either of the two filters. A 'Both Filters' variable was assigned the value 1 if a record was potentially valid for the 2010 Census AC operation and was valid for the 2009 ACS filter; otherwise, it was assigned the value 0. Thus, the comparison, or excluded, category consisted of records that were valid for only one of the two filters.

Also in Section 3.2.1.1 Universe Description, prior research (Holland, 2009) indicated a potential link between imputation type and geocoding accuracy. Imputation type was coded into two binomial variables. If the imputation type was a Point imputation, the Point Imputation variable was coded as the value 1; otherwise, it was assigned the value 0. If the imputation type was a Best Address imputation, the Best Address variable was coded as the value 1; otherwise, it was assigned the value 0. The other address and best ZIP imputations make up the comparison, or excluded, category.

3.2.2. Administrative Records

For this research topic, the DSSD focused on AR data from the Statistical Administrative Records System (StARS), to evaluate how well AR data can update the MTdb. StARS is composed of AR data collected from other federal agencies, including the Internal Revenue Service (IRS), Centers for Medicare and Medicaid Services, Department of Housing and Urban Development, Indian Health Service, and Selective Service System.

As a preliminary step in determining the usefulness of StARS in updating the MTdb, all unique MTdb location and mailing address records in the pre-AC and post-AC files[22] were assembled. Valid addresses for this analysis fell into three basic address types: city-style addresses, rural route addresses, and Post Office (PO) box addresses. Addresses were required to have a non-blank address type and a ZIP Code of length five in order to be included in the input file for StARS matching. About 225 million unique addresses fit these criteria. These addresses came from 184 million MAF records with unique MAF IDentifiers (MAFIDs). In order to maximize the AR-to-census matching, all unique addresses for each MAF record selected were sent to the Data Integration Division (DID) for matching.

Nearly 3 million MAF records had insufficient address data for the matching process. Due to these missing 2,985,007 address records, the distribution of 2010 Census AC Actions for this research topic is different from the total U.S. counts reflected in the 2010 Census AC Assessment. Nearly 2 million of these records had 2010 Census AC actions, including 837,994 2010 Census AC delete actions and 495,784 2010 Census AC duplicate actions.

These addresses were sorted to remove duplicate addresses for each record[23], then passed to the DID to be matched to the 2007 and 2008 StARS Master Housing Files (MHF) per DSSD specifications (Tomaszewski, 2010b). The DID produced a one-to-many address match, matching multiple MAF addresses to each StARS record, and also determined the best MTdb address match for each StARS record. The first matching pass produced successful unit-level matches, and was followed by a Basic Street Address (BSA)[24] level match on the remaining unmatched StARS records.

The DSSD unduplicated the results of the unit-level best MTdb address match by MAFID before merging to the 2010 Census Combo file and the 2010 CUF. The DSSD used the 2010 Census AC results to evaluate whether the StARS database could be of use in updating the MTdb.

The DSSD looked at the existence of MTdb addresses in the StARS 2007 and 2008 MHFs and followed these records through the 2010 Census AC operation to their final census status.

3.2.3. Local GIS Coordinates

In order to evaluate how local GIS coordinates could improve address list maintenance, DSSD examined MTdb records flagged with SCEMA (described in Section 2.1) action codes. Change SCEMA actions indicated corrected geocodes, while add SCEMA actions indicated new geocodes. The SCEMA program used these local GIS coordinates to add addresses to the MTdb,

---

[22] From Section 3.2.1, the 2010 Census Combo file was not available when this input file was created.

[23] Each MAFID could have up to four different addresses: pre-AC mailing address, pre-AC location address, post-AC mailing address, and post-AC location address. Of the up to four addresses, only the unique ones were passed to DID for matching.

[24] The BSA is the address without any unit information, such as apartment numbers. An example address, 101 Main Street, Apartment 1A, has the BSA of 101 Main Street.

identify geocodes for ungeocoded addresses, and to verify or correct geocodes on existing, geocoded addresses in the MTdb (Dixon and Tomaszewski, 2007).

For this research topic (Local GIS coordinates), the DSSD used the January 2009 ACS supplemental sample MAFX to identify the appropriate universe of address records. The MAFX did not record the verification of geocodes; it only recorded add or change actions, so the DSSD examined the GIS-updated addresses flagged as add actions (new geocodes) and change actions (corrected geocodes).

For this research, the DSSD extracted additional data from the 2010 Census Combo file and the 2010 CUF (e.g.[25], 2010 Census AC action codes and other variables of interest). The DSSD used 2010 Census AC actions from the 2010 Census Combo file to evaluate the accuracy of the GIS information. As a supplemental measure, the DSSD also used the 2010 CUF to determine the final census status of these records.

3.2.4. American Community Survey Time of Interview Actions

The ACS selected a sample of approximately 2.9 million addresses (combination of housing units and group quarters) annually from 2005 to 2010. According to the ACS Design and Methodology (Bureau of the Census, 2009a), which discussed the 2007 ACS, approximately 47 percent, or about 1.4 million, were interviewed by mail; another 10 percent, or about 0.3 million, were interviewed by phone; and the remainder (43 percent or about 1.2 million) were sampled for CAPI interviews at various rates (2 in 3, 1 in 2, 2 in 5, or 1 in 3) depending on whether the address was mailable or not, and the expected rate of interview completion. CAPI workloads average over 40 thousand cases a month, which results in an annual workload of over 480 thousand cases a year, or about 17 percent of the annual sample. Approximately 80 percent of CAPI interviews require a personal visit, which is about 384 thousand, or about 13 percent of the annual sample. The ACS TOI action codes examined in this report cover all ACS TOI actions recorded on the MTdb up through January 2009, including ACS TOI actions from as far back as fall 2004.

In order to evaluate how ACS TOI data could be used to improve address list maintenance, the DSSD examined MTdb records flagged with ACS TOI action codes. For the ACS TOI actions research topic, the DSSD used the January 2009 ACS supplemental sample MAFX to identify the following records:

- 390,867 records with ACS TOI change actions (0.27 percent of the 2010 Census AC eligible workload),
- 93,760 records with ACS TOI delete actions (0.07 percent of the 2010 Census AC eligible workload), and
- 20,749 with ACS TOI nonresidential actions (0.01 percent of the 2010 Census AC eligible workload).

[25] Exempli gratia is a Latin phrase meaning ‘for example’ that is commonly abbreviated as ‘e.g.’

The DSSD extracted those records from the 2010 Census Combo file and the CUF, selecting the AC action codes and other variables of interest. The DSSD used AC actions from the 2010 Census Combo file to evaluate the accuracy of the ACS TOI actions. As a supplemental measure, the DSSD also used the CUF to determine the final census status of these records.

4. Limitations

- Puerto Rico was out of scope for all research topics in this evaluation.
- Since this evaluation used a national sample to evaluate the geocode imputations, the estimates are subject to sampling error. Estimates based on a sample can differ from the results obtained from surveying all Housing Units (HUs) and standard errors provide an indication of the minimum possible error present in the estimates. The sample was selected from the combined universe of the potentially-valid for 2010 Census AC ungeocoded records and the January 2009 ACS-valid ungeocoded records, not from all HUs on the MTdb. The standard errors provided with the results primarily reflect variations due to sampling, and do not generally account for nonsampling errors. Nonsampling error describes errors that are not a function of selecting a sample and are difficult to quantify. Nonsampling errors can include errors during data collection and data processing.
- The results file returned to DSSD from the supplemental field operation to evaluate the geocode imputations did not contain 117 of the 6,459 address records originally identified for the national sample.
- The supplemental field operation to evaluate the geocode imputations was conducted over two years after the original 2010 Census AC dependent listing. Changes on the ground, particularly newly constructed or demolished units, may have affected results.
- Ground truth measures depend upon the quality and accuracy of data received from the 2010 Census AC operation and later operations. AC results and CUF status were an approximation of ground truth and may not have corresponded perfectly to actual conditions on the ground. Errors in data collection and processing may have existed in the data and influenced analysis.
- Comparisons were made to the 2010 Census AC operation, excluding Puerto Rico, available in the appendix of the 2010 Census AC Assessment. Comparisons also involved the 2010 Census AC filter in this report. The records sent to AC included Local Update of Census Addresses (LUCA) records, which would not be included in a non-decennial filter. Comparisons to the 2010 Census AC universe and workload differ from the otherwise valid MTdb records by the addition of LUCA records, which may bias such comparisons.

5. Results

5.1. Geocode Imputations

How can we use geocode imputations to improve address list maintenance?
How accurate are imputed geocodes for ungeocoded MTdb records?

The GEO has been studying a software package (called GeoStan), which they licensed for various data cleansing and enhancement features, including its ability to impute geocodes for addresses. The GEO has not used this software to update geocode information on the MTdb

because the quality of the geocodes it provides has not been fully evaluated. The software uses different levels of information to compute geocode imputations. For some addresses, the software uses the entire address (most precise), while for others the software may only have enough information to base the geocode on the ZIP Code (least precise). The software package also contains information connecting some addresses to latitude and longitude coordinate points and house number ranges (associated with streets).

Using imputed geocodes may reduce the risk of FRs or Enumerators missing addresses during future census and survey activities. This is advantageous, since missing an address during listing or other operations can directly contribute to gross undercoverage in a census or survey. During census listings and operations, which are usually block based, correctly imputed addresses will save field costs since these addresses will not need to be added.

Disadvantages of using geocode imputations include the risk of adding mis-geocoded records or duplication to the MTdb, which may lead to increased field costs or tabulation errors. Duplication could occur when a newly geocoded city-style addresses identifies a pre-existing non-city-style MTdb address, and there is no link between the two addresses on the MTdb. This would increase field costs as a link need to be created between the two addresses, generally by marking one address as a duplicate. Mis-geocoded records increase field costs, as the address has to be deleted from the incorrectly imputed block and added to the correct block. If an address is mis-geocoded to the wrong side of a legal boundary and the error is not identified, a tabulation error may occur.

These disadvantages are less concerning for ACS, as ACS FRs are asked to search the entire county for city-style sample addresses[26] that are ungeocoded or mis-geocoded. In this survey, when the imputed block is in close proximity to the correct block, having the imputed block often provides the ACS FR a good starting point to search for the in-sample address. Additionally, in some counties, all ungeocoded units are invalid for ACS due to the risk that they duplicate existing MAF records that do not have city-style addresses. Knowing the geocode or approximate location allows ACS to better filter addresses to distinguish new growth addresses from duplicates of existing non-city-style records. In both cases, additional newly geocoded records, which are correct or close to the correct block, offer an improved sampling frame for the ACS.

From Section 3.2.1, 6,440,043 ungeocoded addresses with imputed geocodes were identified for this research topic from the combined universe of 2009 ACS-valid and otherwise-valid 2010 Census AC ungeocoded, city-style records.

[26] All ungeocoded addresses are city-style addresses. ACS FRs search in the neighborhood/general vicinity for rural route box sample addresses, not the entire county. For sample addresses that are a location/physical description, ACS FRs only search in the block and across the street (due to the vagueness or commonality of the available address information).

Table 4 shows the quantity of address records with imputed geocodes, stratified by imputation type and 2010 Census AC block agreement.

Table 4. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: Geocode Imputation Universe by Imputation Type and 2010 Census AC Block Agreement

Geocode Imputation Types	Address Records Found in AC			Records Not Found in AC	Total
	Found in Imputed Block	Found in Different Block	Total Found in AC		
Point*	1,033,090	31,611	1,064,701	791,945	1,856,646
	(55.6%)	(1.7%)	(57.3%)	(42.7%)	(100.0%)
Best Address**	760,164	177,451	937,615	1,204,270	2,141,885
	(35.5%)	(8.3%)	(43.8%)	(56.2%)	(100.0%)
Other Address; Best ZIP Code***	139,170	108,620	247,790	440,635	688,425
	(20.2%)	(15.8%)	(36.0%)	(64.0%)	(100.0%)
Other ZIP Code****	72,328	521,994	594,322	1,158,765	1,753,087
	(4.1%)	(29.8%)	(33.9%)	(66.1%)	(100.0%)
Total	2,004,752	839,676	2,844,428	3,595,615	6,440,043
	(31.1%)	(13.0%)	(44.2%)	(55.8%)	(100.0%)

Sources: Pre-AC file, Post-AC file, January 2009 ACS MAFX, and GeoStan Imputations File.
Note: Percentages may not sum to 100 due to rounding.
Note: Geostan has internal codes describing the imputation type that were useful in this analysis.
*: Point imputations group together the location codes beginning with AP.
**: Best address imputations group together the location codes of AS0 and AS1.
***: Other address imputations group together the location codes beginning with AI and the remaining location codes beginning with AS. Best ZIP imputations group together the location codes beginning with ZB9.
****: Other ZIP imputations group together the remaining location codes beginning with Z.

The geocode imputation types shown in Table 4 are ordered by 2010 Census AC block agreement. For this analysis, the 2010 Census AC block agreement was determined using the MAF updates GEO applied to the MTdb after the 2010 Census AC operation. Point imputations were both the most likely to exist on the ground (57.3 percent) and the most likely to be found in the imputed block (55.6 percent), while Other ZIP Code imputations were the least likely to exist on the ground (33.9 percent) and the least likely to be found in the imputed block (4.1 percent).

Table 5 shows the quantity of address records with imputed geocodes in this research topic that were found in the 2010 Census AC operation, and their concordance with the 2010 Census geocode.

Table 5. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: Accuracy of Geocode Imputations for Records Found in the 2010 Census AC Operation

Geocode Imputation Types	Address Records where the 2010 Census AC Geocode and the Imputed Geocode Agree		Total Address Records Found in 2010 Census AC	
	Count	Percent of Total	Count	Percent
Point	1,033,090	97.0	1,064,701	100.0
Best Address	760,164	81.1	937,615	100.0
Other Address; Best ZIP Code	139,170	56.2	247,790	100.0
Other ZIP Code	72,328	12.2	594,322	100.0
Total	2,004,752	70.5	2,844,428	100.0

Sources: Pre-AC file, Post-AC file, January 2009 ACS MAFX, and GeoStan Imputations File.
Note: Percentages may not sum to 100 due to rounding.

For this universe, Point geocode imputations were found to be extremely accurate; approximately 97 percent of Point geocode imputations agreed with the 2010 Census AC geocode. This was not unexpected since latitude and longitude coordinate points used to impute Point geocodes usually corresponded to the center of a parcel of land (usually designated/divided based on property ownership). These imputed values also could have been centered on a driveway, structure, front door, or property boundary.

While the accuracy variations were most dramatic by imputation type, there was also regional variation. Nine states (Nevada, Delaware, Connecticut, California, Hawaii, Utah, Oregon, Colorado, and Tennessee) and the District of Columbia had greater than 80 percent accuracy among ungeocoded records after the 2010 Census AC operation. Five states (Mississippi, South Dakota, Kentucky, Wyoming, and West Virginia) had overall accuracy rates under 50 percent. State accuracy statistics are located in the Appendix.

5.1.1. 2010 Census Status

In addition to comparing the geocode imputation values to the 2010 Census AC operation, the DSSD also investigated the 2010 Census final status on the CUF for the universe in this research topic. Table 6 shows a strong correlation between whether a record was found in the 2010 Census AC operation and whether a record was valid on the CUF.

Table 6. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: Geocode Imputation Universe by 2010 Census AC Status and Final 2010 Census Status

2010 Census AC Status	In Census		Not In Census		Total	
	Count	Percent	Count	Percent	Count	Percent
In AC	2,752,823	42.7	91,605	1.4	2,844,428	44.2
Not In AC	583,631	9.1	3,011,984	46.8	3,595,615	55.8
Total	3,336,454	51.8	3,103,589	48.2	6,440,043	100.0

Sources: Pre-AC file, Post-AC file, January 2009 ACS MAFX, 2010 Census Unedited File, and GeoStan Imputations File.
Note: Percentages may not sum to 100 due to rounding.

Nearly 47 percent of the 6.4 million ungeocoded universe was neither found during the 2010 Census AC operation nor valid for the 2010 Census, so, of the previously ungeocoded records

that were not found in the 2010 Census AC operation a little over 3.0 million, or nearly 84 percent, were also not valid records for the 2010 Census. The next largest portion of the ungeocoded universe, nearly 43 percent, was in both the 2010 Census AC operation and the 2010 Census. Or, in other words, nearly 97 percent of the previously ungeocoded records that were found in the 2010 Census AC operation were included in the 2010 Census.

Table 7 shows the final census status for the ungeocoded records found in the 2010 Census AC operation by geocode imputation type. The 2010 Census AC accuracy trend among geocode imputation types (seen in Table 4 and Table 5) continued for final 2010 Census accuracy.

Table 7. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: Final 2010 Census Status of Geocode Imputations for Records Found in the 2010 Census AC Operation

Geocode Imputation Types	In Census		Total Address Records Found in 2010 Census AC
	Count	Percent of Total	Count
Point	1,046,915	98	1,064,701
Best Address	900,431	96	937,615
Other Address; Best ZIP Code	238,693	96	247,790
Other ZIP Code	566,784	95	594,322
Total	2,752,823	97	2,844,428

Sources: Pre-AC file, Post-AC file, January 2009 ACS MAFX, 2010 Census Unedited File, and GeoStan Imputations File.
Note: Percentages may not sum to 100 due to rounding.

The Point imputations had the highest final 2010 Census presence, while Other ZIP Code imputations had the lowest final inclusion rate in the 2010 Census. Overall, the previously ungeocoded records that were found in the 2010 Census AC operation were highly likely to be part of the 2010 Census.

5.1.2. Supplemental Field Operation

Since more than 55 percent of the address records in the ungeocoded universe for this research topic were not found in the 2010 Census AC operation, it was necessary to conduct a supplemental field operation to determine the accuracy of this subset. In the supplemental field operation, the listers searched for specific addresses and returned the appropriate geocode. They were instructed to return a geocode even if the address was not built yet but the lot corresponding to an address was identifiable[27]. Table 8 shows the results of the supplemental field operation by geocode imputation type. Addresses that were not found were grouped with addresses that were imputed incorrectly and addresses found on the ground in a different block.

[27] This is different from the AC procedures, as the purpose of this research was to determine if the imputed geocode was correct.

Table 8 shows that the supplemental field operation results are very similar to the results in Table 5. In the supplemental field operation, Point geocode imputations exhibited the highest accuracy at approximately 91 percent, followed by the Best Address geocode imputations at about 73 percent accuracy. These types of geocode imputations were both highly accurate and likely to exist on the ground. The Other Address imputations and Best ZIP Code imputations were found by the supplemental field operation in the imputed block just over half the time. This indicates that addresses with Point imputations or Best Address imputations are both likely to be correctly located and more likely to have existed on the ground than addresses with Other Address imputations or Best ZIP Code imputations.

Table 8. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: Geocode Imputation Supplemental Field Operation Results – Sample Sizes, Weighted Counts, and Weighted Percentages

Geocode Imputation Type	Found Correct		Found Incorrect or Not Found		Standard Deviation of Weighted Frequency	Standard Error of Weighted Percent
	Sample Size	Weighted Frequency (Percent)	Sample Size	Weighted Frequency (Percent)		
Point	841	1,683,002 (90.6%)	306	173,644 (9.4%)	16,068	0.8655
Best Address	1,232	1,571,666 (73.4%)	1,300	570,219 (26.6%)	18,925	0.8836
Other Address; Best ZIP Code	957	351,403 (51.0%)	1,796	337,022 (49.0%)	6,731	0.9777
Total	3,030	3,606,070 (76.9%)	3,402	1,080,886 (23.1%)	25,722	0.5488

Sources: Pre-AC file, Post-AC file, January 2009 ACS MAFX, and GeoStan Imputations File.

5.1.3. Logistic Regression Analysis

Variables such as geography and multi-unit status may affect geocoding accuracy. In order to show the net effect of geocode imputation type while accounting for other variables that significantly affect geocoding accuracy, we developed a logistic regression model. A logistic regression model is an easily understandable way to compare the relative effects of several different variables on an outcome variable, in this case the accuracy of geocoding.

The logistic regression analysis shown in Table 9 indicates that the Point imputation method was the most powerful indicator of accuracy of the imputation outcome, while controlling for region, address filters, and multi-unit status.

Table 9. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: Regression Analysis on the National Sample of Geocode Imputation Supplemental Field Operation

Variable	Estimated Coefficient	Wald Chi-Square	Chi-Square Probability	Odds Ratio Estimate	Wald 95 Percent Confidence Limits
Intercept	-0.76	40.88	<.0001	-	-
Northeast	0.13	0.72	0.3966	1.138	0.844 - 1.533
South	0.36	10.74	0.0010	1.439	1.158 - 1.789
West	0.37	6.31	0.0120	1.452	1.085 - 1.942
Multi-unit	0.23	4.02	0.0449	1.253	1.005 - 1.561
Both Filters	0.73	65.52	<.0001	2.078	1.741 - 2.481
Point	2.10	326.47	<.0001	8.188	6.518 -10.286
Best Address	0.96	238.85	<.0001	2.616	2.316 - 2.955

Sources: Pre-AC file, Post-AC file, January 2009 ACS MAFX, and GeoStan Imputations File.
Note: The weighted number of records used in this table is 4,986,956.

The model fit very well overall, with a max-rescaled R^2 value[28] approaching 1. Of the region variables, the Northeast was not statistically significant at an alpha of 0.0500. The multi-unit variable was marginally significant with a p of $0.0449 < 0.0500$. However, none of the regional nor the multi-unit variables made a meaningful contribution to the odds of an accurate geocode prediction, compared to the rest of the variables in the equation. The number of records passing both address filters (2010 Census AC and 2009 ACS) was statistically significant at $p < 0.0001$, with an increase in the odds of an accurate geocode prediction of just over 200 percent.

The two geocode imputation types were also statistically significant at $p < 0.0001$. The Point geocode imputation type was by far the superior imputation type with an odds ratio of over 8, an increase in the odds of an accurate geocode prediction of over 800 percent.

5.2. Administrative Records

How can we use AR to improve address list maintenance?
Do AR provide intelligence for adding new records, confirming existing MTdb records, and/or identifying deletes or vacant units?

For the AR topic, DSSD evaluated the contribution potential of StARS, including but not limited to:

- how many records were added in the 2010 Census AC operation that previously existed in StARS (potential new adds),
- how many records were present in both MTdb and StARS (confirmation of existing records), and

[28] This is a pseudo-R^2 value and only approximates the meaning of an ordinary least squares R^2 value.

- how many records were deleted in the 2010 Census AC operation that were absent from StARS (intelligence that could be used to help identify and/or confirm nonexistent or vacant units).

StARS is a potential source of information that may assist in address list maintenance. Stakeholders have expressed interest in the utility of AR data (beyond the DSF) in updating the MTdb. While internal and external policy hurdles persist for using IRS data, this evaluation did not address these concerns; but rather focused on the quantitative value of using this data source to maintain the MTdb.

5.2.1. AR and MTdb Universes Eligible for Matching

For this research topic, approximately 184 million address records with complete address types and ZIP Codes were identified from the MTdb. Up to four unique addresses for each of these 184 million address records were created using the pre- and post- 2010 Census AC operation mailing and location addresses. This resulted in approximately 225 million unique addresses (listed in Table 10 by address type) being eligible for matching to the StARS 2007 and StARS 2008 databases. The StARS 2007 MHF contains 152,203,709 addresses and the StARS 2008 MHF contains 154,343,213 addresses. For this research topic, the DSSD focused on how StARS matched the MTdb at a unit level.

Table 10 indicates that city-style addresses were the most common type of MTdb address eligible for StARS matching, at nearly 97 percent of all unique MTdb addresses.

Table 10. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: MTdb Addresses Eligible for StARS Matching by Address Type

Address Type	Unique MTdb Addresses Eligible for StARS Matching	Percent
City-Style	217,927,468	96.9
Multi-Unit	73,590,084	32.7
Single Unit	144,337,384	64.2
Non-City-Style	7,000,246	3.1
Post Office Box	1,131,408	0.5
Rural Route	2,936,609	1.3
Incomplete	2,932,229	1.3
Total	224,927,714	100.0

Source: Pre-AC file and Post-AC file.

Note: Percentages may not sum to 100 due to rounding.

Table 11 shows that non-city-style StARS addresses often did not match at a unit level to MTdb addresses. Over one-third of the nonmatches and duplicate matches (38 percent) were non-city-style addresses. From Table 10, there were only one million PO Box MTdb addresses eligible for matching. Note that the majority of the StARS non-city-style nonmatches (over 11 million on each of the two StARS databases) were PO Box addresses. DSSD was also interested in investigating if addresses with blank unit designations (BUD) were an issue in matching, but the data available for this evaluation did not contain an indicator of BUD addresses of sufficient accuracy to address this issue.

Table 11. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: 2007-8 StARS Unit-Level Nonmatches and Duplicates by Address Types

Address Type	2007 StARS	Percent	2008 StARS	Percent
City-Style	24,613,478	62	25,010,151	62
Multi-Unit	11,308,708	28	11,658,540	29
Single Unit	13,304,770	34	13,351,611	33
Non-City-Style	15,081,187	38	15,372,803	38
Post Office Box	11,696,995	29	11,719,022	29
Rural Route	1,656,710	4	1,500,424	4
Incomplete	1,727,482	4	2,153,357	5
Total	39,694,665	100	40,382,954	100

Source: Pre-AC file, Post-AC file, 2010 Census Combo file, and StARS 2007-8 files.

Note: Percentages may not sum to 100 due to rounding.

Table 12 shows that city-style addresses were the most common type of StARS address that matched at a unit level to the MTdb. Few non-city-style StARS addresses matched to the MTdb universe of addresses for this research topic.

Table 12. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: Address Types of Unduplicated Unit Level 2007-8 StARS Matches

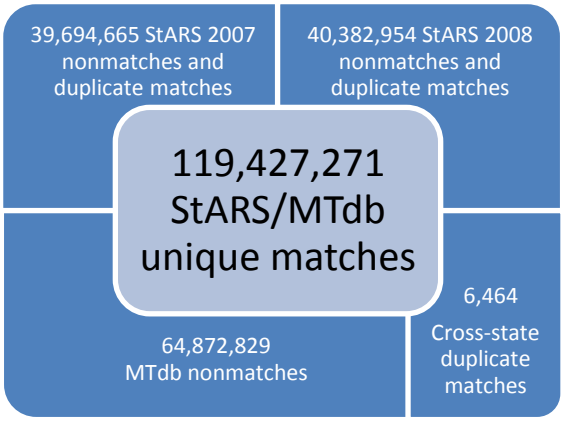
Address Type	2007 StARS	Percent	2008 StARS	Percent	Both StARS	Percent
City-Style	112,506,196	100	113,957,379	100	119,424,157	100
CS – Multi-Unit	24,311,677	23	24,923,731	22	27,526,580	23
CS – Single Unit	88,194,519	77	89,033,648	78	91,897,577	77
Non-City-Style	2,848	0	2,880	0	3,114	0
NCS – Post Office Box	0	0	460	0	334	0
NCS – Rural Route	5	0	52	0	39	0
NCS – Incomplete	2,843	0	2,368	0	2,741	0
Total Matching Records	112,509,044	100	113,960,259	100	119,427,271	100

Source: Pre-AC file, Post-AC file, 2010 Census Combo file, and StARS 2007-8 files.

Note: Percentages may not sum to 100 due to rounding.

Each StARS address was matched at a unit level to the best matching MTdb address (one to four addresses for each of the 184,300,100 MTdb records eligible for matching). In some instances, different StARS address records matched to the same MTdb address record. Unduplication of the StARS/MTdb matches by MAFID resulted in 112,509,044 StARS 2007 address records and 113,960,259 StARS 2008 address records that matched to unique MTdb records, as can be seen in Table 12. Matching the MAFIDs between these two years of StARS yielded a total of 119,427,271 unique StARS/MTdb matches, as seen in Table 12 and in the accompanying Figure 3.

Figure 3. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: Summary of StARS/MTdb Matches, Nonmatches, and Duplicates



Source: Pre-AC file, Post-AC file, 2010 Census Combo file, and StARS 2007-8 files.
Note that the above figure is not to scale.

5.2.2. AR Matches by 2010 Census AC Filter and Actions

The StARS address records that matched to the MTdb were unduplicated by MAFID and merged with the 2010 Census Combo file and the 2010 CUF. This match provided information on the 2010 Census AC filter, AC actions, and final census status. (Note that StARS records were not matched directly to the 2010 CUF, so this analysis did not cover post-AC adds during enumeration operations that may have matched to StARS records.) The 2010 Census AC actions and final 2010 census status were derived from the 2010 Census. Similar decennial census information will not be available for another ten years.

However, the 2010 Census AC filter is a collection of rules to identify valid MTdb addresses that the Census Bureau could recreate every year if desirable. StARS data could also be assembled every year, providing useful updates between censuses. Table 13 shows that nearly 60 percent of the address records valid for the 2010 Census AC filter matched to StARS.

Table 13. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: MTdb/StARS Match Status by 2010 Census AC Filter

2010 Census AC Filter	StARS Match	StARS Nonmatch	Total
In AC Filter	109,031,356	32,734,928	141,766,284
	(59.2%)	(17.8%)	
Not in AC Filter	10,395,915	32,137,900	42,533,815
	(5.6%)	(17.4%)	
Total	119,427,271	64,872,828	184,300,099

Source: Pre-AC file, Post-AC file, 2010 Census Combo file, and StARS 2007-8 files.
Note: Percentages may not Sum to 100 due to rounding.

Table 14 shows the distribution of the 2010 Census AC action codes for the MTdb/StARS matches and the MTdb/StARS nonmatches. MTdb/StARS matches existed for 65 percent (119,427,271/184,300,099) of the address list. If records without 2010 Census AC action codes are excluded to approximate the 2010 Census AC workload, then about 79 percent (119,427,271/153,171,332) of the 2010 Census AC workload had a StARS address match.

Table 14. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: MTdb/StARS Match Status by 2010 Census AC Action Code

2010 Census AC Action Code	MTdb/StARS Match	MTdb/StARS Nonmatch	MTdb Total	StARS Valid to Invalid Ratio
Adds	3,006,304	6,901,473	9,907,777	0.44
New Adds	265,163	5,571,084	5,836,247	0.05
Matched Adds	2,741,141	1,330,389	4,071,530	2.06
Verify	85,909,962	11,601,366	97,511,328	7.41
Change	12,639,561	5,805,248	18,444,809	2.18
Move	4,813,958	631,709	5,445,667	7.62
Delete (verified)	3,832,337	10,859,392	14,691,729	0.35
Single Delete	371	345	716	1.08
Duplicate	1,021,651	2,451,060	3,472,711	0.42
Nonresidential	343,450	857,993	1,201,443	0.40
Uninhabitable	170,351	352,895	523,246	0.48
Invalid/No Action	7,689,326	25,411,347	33,100,673	0.30
Total records	119,427,271	64,872,828	184,300,099	1.84

Source: Pre-AC file, Post-AC file, 2010 Census Combo file, StARS 2007-8 files, and 2010 Census Unedited File.

The last column in Table 14 is the StARS Valid to Invalid Ratio. This is the number of address records that were valid in StARS (the MTdb/StARS Match column) divided by the number of records that were invalid in StARS (the MTdb/StARS Nonmatch column) for each 2010 Census AC action code. The positive 2010 Census AC actions were more likely to match to records in StARS, while the negative 2010 Census AC actions were less likely to match to records in StARS. To investigate this trend further, first records with positive actions such as add, change, verify, or move were examined; then records with negative actions.

5.2.3. AR Matching Status for 2010 Census AC Adds through the CUF

Overall, 30 percent (3,006,304) of the 9,907,777 AC adds existed in StARS (vintage 2007 or 2008) before the 2010 Census AC operation. Those StARS-matched 2010 Census AC adds were primarily adds that matched pre-existing MAF records (2,741,141, or 67 percent of the 4,071,530 matched adds), with a small number of new AC adds (265,163, or less than five percent of the 5,836,247 new AC adds). The majority of the 3,006,304 StARS-matched 2010 Census AC adds were also valid for the 2010 Census (2,938,624, or 98 percent), with a ratio of 43 valid records in 2010 Census for every record not valid in the 2010 Census. The balance of 2010 Census AC adds, about 70 percent (6,901,473), did not exist in StARS prior to the 2010 Census AC operation. The majority of those 6,901,473 non-StARS adds were valid 2010 Census records (5,627,641, or 82 percent), while a smaller number of the non-StARS adds (1,273,832, or 18 percent) were not in the final 2010 Census. The non-StARS adds had a 2010 Census valid to invalid ratio of 4.4:1.

To examine the differences between new 2010 Census AC adds and matched 2010 Census AC adds, the DSSD tracked these universes through the assignment of final 2010 Census validity, shown in Table 15 and Table 16. Table 15 shows that 265,163, or 4.5 percent, of the 5,836,247 new 2010 Census AC adds existed in the 2007 or 2008 StARS before the 2010 Census AC operation.

Table 15. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: 2010 Census AC New Adds by StARS Match Status and 2010 CUF Validity

2010 Census AC Action Code	StARS Match Status	CUF Validity	Count	Percent	CUF Valid to Invalid Ratio
New Adds.....	Nonmatch	Invalid	1,151,502	19.7	3.84
	Nonmatch	Valid	4,419,582	75.7	
	Match	Invalid	24,703	0.4	9.73
	Match	Valid	240,460	4.1	
Total			5,836,247	100.0	

Source: Pre-AC file, Post-AC file, 2010 Census Combo file, StARS 2007-8 files, and 2010 Census Unedited File.
Note: Percentages may not sum to 100 due to rounding.

The remaining 5,571,084, or 95.5 percent, of the new 2010 Census AC adds did not exist in StARS before the 2010 Census AC operation. New add address records are often new construction, and addresses must be occupied by tax filers before appearing in StARS via an IRS source (99 percent of StARS records that matched had an IRS source for the address). This helps explain the high StARS nonmatch rate for this universe.

Interestingly, 1,151,502/5,836,247, or 19.7 percent, of the new 2010 Census AC adds both did not exist in StARS before the 2010 Census AC operation and were not considered valid for the 2010 Census.

Table 16 shows that two-thirds (2,741,141, or 67 percent) of 2010 Census AC adds that matched to pre-existing MTdb records also existed in StARS before the 2010 Census AC operation. Nearly all of these 2,741,141 matched adds records in StARS (2,698,164, or 98 percent) were also valid for the 2010 Census, with a ratio of 63 valid records in the 2010 Census for every record not valid in the 2010 Census.

Table 16. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: 2010 Census AC Matched Adds by StARS Match Status and 2010 CUF Validity

2010 Census AC Action Code	StARS Match Status	CUF Validity	Count	Percent	CUF Valid to Invalid Ratio
Matched Adds.....	Nonmatch	Invalid	122,330	3.0	9.88
	Nonmatch	Valid	1,208,059	29.7	
	Match	Invalid	42,977	1.1	62.78
	Match	Valid	2,698,164	66.3	
Total			4,071,530	100.0	

Source: Pre-AC file, Post-AC file, 2010 Census Combo file, StARS 2007-8 files, and 2010 Census Unedited File.
Note: Percentages may not sum to 100 due to rounding.

The remaining one-third (1,330,389, or 33 percent) of matched 2010 Census AC adds, which did not exist in StARS, also showed a high validity percentage (91 percent) for the 2010 Census with a ratio of 10 valid records in the 2010 Census for every record not valid in the 2010 Census.

5.2.4. AR Matching Status for Positive (Non-Add) 2010 Census Actions through the CUF

Table 17 shows that over two-thirds of records with 2010 Census AC change actions (12,639,561/18,444,809, or 69 percent) matched to 2007 StARS or 2008 StARS address records.

Table 17. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: 2010 Census AC Change Actions by StARS Match Status and 2010 CUF Validity

2010 Census AC Action Code	StARS Match Status	CUF Validity	Count	Percent	CUF Valid to Invalid Ratio
Change.....	Nonmatch	Invalid	635,501	3.4	8.13
	Nonmatch	Valid	5,169,747	28.0	
	Match	Invalid	256,660	1.4	48.25
	Match	Valid	12,382,901	67.1	
Total			18,444,809	100.0	

Source: Pre-AC file, Post-AC file, 2010 Census Combo file, StARS 2007-8 files, and 2010 Census Unedited File.
Note: Percentages may not sum to 100 due to rounding.

The distribution for the 18,444,809 records with 2010 Census AC change actions (address changes) in StARS and in the 2010 CUF was similar to the distribution for matched 2010 Census AC adds. Two-thirds of records with change actions (12,382,901, or 67 percent) existed in StARS prior to the 2010 Census AC operation and were valid records for the 2010 Census, while nearly one-third of records with change actions (5,169,747, or 28 percent) did not exist in StARS prior to the 2010 Census AC operation but were valid records for the 2010 Census.

Table 18 shows that nearly nine-tenths (85,909,962, or 88 percent) of the 97,511,328 address records with verify actions in the 2010 Census AC operation matched to 2007 or 2008 StARS address records. Nearly all of these verify address records in StARS (85,251,065/85,909,962, or 99 percent) were valid records in the 2010 Census, with a ratio of 129 valid records in the 2010 Census for every record not valid in the 2010 Census.

Table 18. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: 2010 Census AC Verify Actions by StARS Match Status and 2010 CUF Validity

2010 Census AC Action Code	StARS Match Status	CUF Validity	Count	Percent	CUF Valid to Invalid Ratio
Verify	Nonmatch	Invalid	1,042,182	1.1	10.13
	Nonmatch	Valid	10,559,184	10.8	
	Match	Invalid	658,897	0.7	129.38
	Match	Valid	85,251,065	87.4	
Total			97,511,328	100.0	

Source: Pre-AC file, Post-AC file, 2010 Census Combo file, StARS 2007-8 files, and 2010 Census Unedited File.
Note: Percentages may not sum to 100 due to rounding.

A smaller number of address records with 2010 Census AC verify actions (10,559,184/97,511,328, or nearly 11 percent) did not exist in StARS prior to the 2010 Census AC operation but were valid records in the 2010 Census, with a ratio of 10 valid records in the 2010 Census for every record not valid in the 2010 Census.

Table 19 shows that nearly nine-tenths (4,813,958/5,445,667, or 88 percent) of address records with 2010 Census AC move actions (where an address was deleted from one census block and added to a different census block) matched to 2007 StARS or 2008 StARS address records. This distribution is similar to that of 2010 Census AC verified address records.

Table 19. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: 2010 Census AC Move Actions by StARS Match Status and 2010 CUF Validity

2010 Census AC Action Code	StARS Match Status	CUF Validity	Count	Percent	CUF Valid to Invalid Ratio
Move	Nonmatch	Invalid	51,834	1.0	11.19
	Nonmatch	Valid	579,875	10.6	
	Match	Invalid	42,958	0.8	111.06
	Match	Valid	4,771,000	87.6	
Total			5,445,667	100.0	

Source: Pre-AC file, Post-AC file, 2010 Census Combo file, StARS 2007-8 files, and 2010 Census Unedited File.
Note: Percentages may not sum to 100 due to rounding.

The distribution for address records with 2010 Census AC move actions in StARS and in the 2010 CUF was similar to the distribution of 2010 Census AC verify address records. Nearly all of the records with 2010 Census AC move actions that were in StARS (4,771,000/4,813,958, or 99 percent) were valid records in the 2010 Census, with a ratio of 111 valid records in the 2010 Census for every record not valid in the 2010 Census. A smaller number of moved records (579,875/5,445,667 or 11 percent) did not exist in StARS prior to the 2010 Census AC operation but were valid records for the 2010 Census, with a ratio of 11 valid records in the 2010 Census for every record not valid in the 2010 Census.

Overall, records with positive 2010 Census AC actions (verify, move, change) were likely to exist in StARS, with a low of 68.5 percent of records with 2010 Census AC change actions matching to StARS, to a high of 88.4 percent of records with 2010 Census AC move actions matching to StARS. Similarly, a high percentage of 2010 Census AC adds (67.3 percent) that matched to pre-existing MTdb address records also existed in StARS. However, the same could not be said for the new 2010 Census AC adds.

5.2.5. AR Matching Status for Negative or No 2010 Census AC Actions through the CUF

In Table 20 through Table 24, address records with negative 2010 Census AC actions (delete, duplicate, nonresidential, uninhabitable) and no 2010 Census AC actions are examined.

Table 20 shows that just over one-fourth of address records with 2010 Census AC delete actions (26.1 percent) matched to a 2007 StARS or 2008 StARS address. The 2010 Census AC deletes in Table 20 include both verified delete actions and single delete actions. (There were only 716 single delete actions in the 2010 Census AC operation.)

Table 20. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: 2010 Census AC Delete Actions by StARS Match Status and 2010 CUF Validity

2010 Census AC Action Code	StARS Match Status	CUF Validity	Count	Percent	CUF Valid to Invalid Ratio
Delete.....	Nonmatch	Invalid	10,489,340	71.4	0.04
	Nonmatch	Valid	370,397	2.5	
	Match	Invalid	3,090,446	21.0	0.24
	Match	Valid	742,262	5.1	
Total			14,692,445	100.0	

Source: Pre-AC file, Post-AC file, 2010 Census Combo file, StARS 2007-8 files, and 2010 Census Unedited File.
Note: Percentages may not sum to 100 due to rounding.

Close to three-fourths of 2010 Census AC deletes (71.4 percent) did not exist in StARS prior to the 2010 Census AC operation and were invalid for the 2010 Census, while 2.5 percent of 2010 Census AC deletes did not exist in StARS prior to the 2010 Census AC operation but were valid for the 2010 Census. About one-fifth of 2010 Census AC deleted records (21.0 percent) did exist in StARS prior to the 2010 Census AC operation but were invalid for the 2010 Census, while 5.1 percent of 2010 Census AC deletes existed in StARS prior to the 2010 Census AC operation and were valid for the 2010 Census.

Ninety-seven percent of the 2010 Census AC deletes that did not match StARS were invalid for the 2010 Census, while 81 percent of 2010 Census AC deletes that matched to StARS were invalid for the 2010 Census. For the 2010 Census AC deletes that did not match to StARS, there were 28 invalid 2010 CUF records for each valid 2010 CUF record (or 0.04 valid 2010 CUF records for every invalid 2010 CUF record). For 2010 Census AC deletes that matched to StARS records, there were 0.24 valid 2010 CUF records for each invalid 2010 CUF record (or 4 invalid 2010 CUF records for each valid 2010 CUF record).

Table 21 shows that less than one-third of address records with 2010 Census AC duplicate actions (29.4 percent) matched to 2007 StARS or 2008 StARS address records.

Table 21. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: 2010 Census AC Duplicate Actions by StARS Match Status and 2010 CUF Validity

2010 Census AC Action Code	StARS Match Status	CUF Validity	Count	Percent	CUF Valid to Invalid Ratio
Duplicate	Nonmatch	Invalid	2,386,256	68.7	0.03
	Nonmatch	Valid	64,804	1.9	
	Match	Invalid	979,834	28.2	0.04
	Match	Valid	41,817	1.2	
Total			3,472,711	100.0	

Source: Pre-AC file, Post-AC file, 2010 Census Combo file, StARS 2007-8 files, and 2010 Census Unedited File.
Note: Percentages may not sum to 100 due to rounding.

Over two-thirds of 2010 Census AC duplicates (68.7 percent) did not exist in StARS prior to the 2010 Census AC operation and were invalid for the 2010 Census, while 1.9 percent of 2010 Census AC duplicates did not exist in StARS prior to the 2010 Census AC operation but were valid for the 2010 Census. For the 2010 Census AC duplicates that did not match to StARS, there were 0.03 valid 2010 CUF address records for every invalid 2010 CUF address record (or 36 2010 Census AC duplicates invalid on the 2010 CUF for each 2010 Census AC duplicate valid on the 2010 CUF). Over one-fourth of 2010 Census AC duplicates (28.2 percent) existed in StARS prior to the 2010 Census AC operation but were invalid for the 2010 Census, while 1.2 percent of 2010 Census AC duplicates existed in StARS prior to the 2010 Census AC operation and were valid for the 2010 Census.

Overall, the majority (71 percent) of both the 2010 Census AC delete and duplicate addresses were not in StARS immediately before the 2010 Census AC operation, and were considered invalid for the 2010 Census. The distribution for 2010 Census AC address records determined nonresidential were similar to the 2010 Census AC address records with delete or duplicate codes.

Table 22 shows us that less than one-third of address records with 2010 Census AC nonresidential actions (28.6 percent) matched to 2007 StARS or 2008 StARS address records.

Table 22. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: 2010 Census AC Nonresidential Actions by StARS Match Status and 2010 CUF Validity

2010 Census AC Action Code	StARS Match Status	CUF Validity	Count	Percent	CUF Valid to Invalid Ratio
Nonresidential	Nonmatch	Invalid	843,884	70.2	0.02
	Nonmatch	Valid	14,109	1.2	
	Match	Invalid	325,690	27.1	0.05
	Match	Valid	17,760	1.5	
Total			1,201,443	100.0	

Source: Pre-AC file, Post-AC file, 2010 Census Combo file, StARS 2007-8 files, and 2010 Census Unedited File.
Note: Percentages may not sum to 100 due to rounding.

Seven-tenths of 2010 Census AC address records with nonresidential action codes (70.2 percent) did not exist in StARS prior to the 2010 Census AC operation and were invalid for the 2010

Census, while 1.2 percent did not exist in StARS prior to the 2010 Census AC operation but were valid for the 2010 Census. About one-fourth of 2010 Census AC address records with nonresidential action codes (27.1 percent) did exist in StARS prior to the 2010 Census AC operation but were invalid for the 2010 Census, while 1.5 percent existed in StARS prior to the 2010 Census AC operation and were valid for the 2010 Census.

Table 23 shows approximately one-third of address records with a 2010 Census AC uninhabitable action code matched to 2007 StARS or 2008 StARS address records.

Table 23. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: 2010 Census AC Uninhabitable Actions by StARS Match Status and 2010 CUF Validity

2010 Census AC Action Code	StARS Match Status	CUF Validity	Count	Percent	CUF Valid to Invalid Ratio
Uninhabitable	Nonmatch	Invalid	192,847	36.9	0.83
	Nonmatch	Valid	160,048	30.6	
	Match	Invalid	77,139	14.7	1.21
	Match	Valid	93,212	17.8	
Total			523,246	100.0	

Source: Pre-AC file, Post-AC file, 2010 Census Combo file, StARS 2007-8 files, and 2010 Census Unedited File.
Note: Percentages may not sum to 100 due to rounding.

Slightly more than one-third of address records with 2010 Census AC uninhabitable action codes (36.9 percent) did not exist in StARS prior to the 2010 Census AC operation and were invalid for the 2010 Census. Just below one-third of 2010 Census AC uninhabitable address records (30.8 percent) did not exist in StARS prior to the 2010 Census AC operation but were valid for the 2010 Census. The remaining one-third of 2010 Census AC uninhabitable addresses existed in StARS prior to the 2010 Census AC operation.

Table 24 shows that over three-fourths of MTdb address records with no 2010 Census AC action code did not match to 2007 StARS or 2008 StARS address records.

Table 24. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: No 2010 Census AC Action by StARS Match Status and 2010 CUF Validity

2010 Census AC Action Code	StARS Match Status	CUF Validity	Count	Percent	CUF Valid to Invalid Ratio
No AC Action.....	Nonmatch	Invalid	24,944,771	75.4	0.02
	Nonmatch	Valid	466,576	1.4	
	Match	Invalid	7,082,630	21.4	0.09
	Match	Valid	606,696	1.8	
Total			33,100,673	100.0	

Source: Pre-AC file, Post-AC file, 2010 Census Combo file, StARS 2007-8 files, and 2010 Census Unedited File.
Note: Percentages may not sum to 100 due to rounding.

Overall, records with a negative 2010 Census AC action code (delete, duplicate, nonresidential, uninhabitable) or no 2010 Census AC action code were not likely to match to StARS, with a low of 67.4 percent of address records with uninhabitable actions missing from StARS, to a high of 76.8 percent of address records with no 2010 Census AC action code absent from StARS. In a way similar to how the presence of an address record in StARS was likely to indicate a positive 2010 Census AC action code (verify, move, change), the absence of an address in StARS is likely to indicate a negative 2010 Census AC action.

5.2.6. AR Matching Status and the 2010 Census AC filter through the CUF

Table 25 summarizes the universe of address records for this research topic, by providing a distribution of 2010 Census AC filter by StARS Match Status by 2010 Census validity. It shows that the majority of address records that both were eligible for the 2010 Census AC operation and existed in StARS were valid for the 2010 Census, with a ratio of 19 valid address records to one invalid record in the 2010 CUF. Similarly, Table 25 shows that the majority of records that both were not eligible for the 2010 Census AC operation and did not exist in StARS were invalid for the 2010 Census, with a ratio of one valid address record to four invalid address records in the 2010 CUF. So, again, there is a strong relationship between 2010 Census AC filter and StARS inclusion with validity in the 2010 Census.

Table 25. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: MTdb Summary - 2010 Census AC Filter Status by StARS Match Status by 2010 CUF Validity

2010 Census AC Filter	StARS Match Status	In 2010 Census	Not In 2010 Census	Total	CUF Valid to Invalid Ratio
In AC	In StARS	103,575,333 (56.2%)	5,456,023 (3.0%)	109,031,356 (59.2%)	18.98
In AC	Not in StARS	17,020,090 (9.2%)	15,714,838 (8.5%)	32,734,928 (17.8%)	1.08
Not in AC	In StARS	3,270,004 (1.8%)	7,125,911 (3.9%)	10,395,915 (5.6%)	0.46
Not in AC	Not in StARS	5,992,291 (3.3%)	26,145,609 (14.2%)	32,137,900 (17.4%)	0.23
Total		129,857,718 (70.5%)	54,442,381 (29.5%)	184,300,099 (100.0%)	

Source: Pre-AC file, Post-AC file, 2010 Census Combo file, StARS 2007-8 files, and 2010 Census Unedited File.
Note: Percentages may not sum to 100 due to rounding.
The yellow and green shaded blocks show where the 2010 Census AC filter correctly predicted the 2010 CUF Validity.
The blue and green shaded blocks show where the StARS match status correctly predicted the 2010 CUF Validity

Table 25 shows that the 2010 Census AC filter was 83.5 percent correct[29] according to the 2010 CUF, while the StARS match status was 80.7 percent correct[30]. The StARS match status and 2010 Census AC filter agree with the CUF for 70.4 percent of address records[31] and disagree

[29] (153,866,943/184,300,099) Correctness of the AC filter is calculated by adding the In AC/In 2010 Census counts of 103,575,333 and 17,020,090 to the Not in AC/Not In 2010 Census counts of 7,125,911 and 26,145,609. In other words, where the AC filter correctly identified records as valid or invalid for the 2010 Census. Adds were not in the AC filter, so they would be part of the Not in AC rows. They only contribute to the correctness of the AC filter if they were later marked as invalid for the 2010 Census. Numbers are highlighted in green and yellow. Green overlaps with the blue of StARS.

[30] (148,705,784/184,300,099) Correctness of the StARS match status is calculated by adding the In StARS/In 2010 Census counts of 103,575,333 and 3,270,004 to the Not in StARS/Not In 2010 Census counts of 15,714,838 and 26,145,609. In other words, where the StARS match Status correctly identified records as valid or invalid for the 2010 Census. Numbers are highlighted in blue and green. Green overlaps with the yellow of the AC filter.

[31] (129,720,942/184,300,099) The table entries are highlighted in green: the In AC/In StARS/In 2010 Census count of 103,575,333 plus the Not in AC/Not in StARS/Not In 2010 Census count of 26,145,609.

with the CUF for 6.2 percent of address records[32]. The StARS match status and 2010 Census AC filter disagree with each other for 23.4 percent of address records[33], with a net difference of 2.8 percent[34] in predicting the correct final 2010 Census status. This indicates a strong similarity in predicting the final 2010 Census status between the StARS match status and 2010 Census AC filter.

Integrating StARS match status into the MTdb may prove useful for future address filters and operational activities. For instance, in a targeting scenario for future operations, address records that matched to StARS and were valid for an address filter could be categorized as valid while records that did not match to StARS and were invalid for an address filter could be considered invalid. In both cases, a field operation would not examine this cases (or blocks, tracks, or other geographical areas with a majority of these records). Records where the StARS match status and address filter disagreed on validity (or geographical areas with a majority of these records) would be targeted for a field operation to determine whether to include these records for future operations. In summary, the relationships shown here are strong –the presence of an address in StARS increases the likelihood of a positive listing outcome and positive final census status; and the absence of an address in StARS increases the likelihood of a negative listing outcome and negative final census status.

5.3. Local Geographic Information Systems Coordinates:

How can we use local GIS coordinates to improve address list maintenance?
Were the geocodes derived from local GIS coordinates accurate?

A match between the January 2009 ACS supplemental sample MAFX and the 2010 Census Combo file yielded 3,707,693 address records with local GIS coordinate updates (adds and corrections), of which 3,536,752 (95.4 percent) were sent to the 2010 Census AC operation.

[32] (11,448,314/184,300,099) The table entries are highlighted in red: the In AC/In StARS/Not In 2010 Census count of 5,456,203 plus the Not in AC/Not in StARS/In 2010 Census count of 26,145,609.

[33] (43,130,843/184,300,099) The table entries are highlighted in blue and yellow: the AC filter's correct, yellow predictions of In AC/Not in StARS/In 2010 Census (17,020,090) and Not in AC/In StARS/Not In 2010 Census (7,125,911), which sum to 24,146,001; and the StARS correct, blue predictions of Not in AC/In StARS/In 2010 Census (3,270,004) and In AC/Not in StARS/Not In 2010 Census (15,714,838), which sum to 18,984,842.

[34] (5,161,159/184,300,099) The difference between the blue entries and the yellow entries is listed in the previous footnote.

Table 26 shows that 95.1 percent of addresses with local GIS coordinate add actions and 95.4 percent of addresses with local GIS coordinate change actions were part of the 2010 Census AC workload.

Table 26. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: GIS Coordinates Address Universe by 2010 Census AC Eligibility and Geocode Type

2010 Census AC Filter	Add Actions (New Geocodes)	Change Actions (Corrected Geocodes)	Total GIS Coordinate based Actions
Eligible	331,643	3,205,109	3,536,752
	(95.1%)	(95.4%)	(95.4%)
Not Eligible	17,266	153,675	170,941
	(4.9%)	(4.6%)	(4.6%)
Total	348,909	3,358,784	3,707,693
	(100.0%)	(100.0%)	(100.0%)

Note: Percentages may not sum to 100 due to rounding.
Source: January 2009 ACS MAFX and the 2010 Census Combo File.

The majority (3,358,784/3,707,693 addresses, or 90.6 percent) of the local GIS coordinate updates were processed as change actions, since they corrected MTdb geocodes. These address records were geocoded on the MTdb, but local GIS coordinate updates geocoded them to different census blocks. Each mis-geocoded address record delivered to the 2010 Census AC operation would have required a delete action in the original block and an add action in the correct block (and for GEO to match the addresses during post-AC processing to identify this error). When accurate, this information corrected geocoding errors before the 2010 Census AC operation, and helped to reduce operational costs.

Address records with local GIS coordinate add actions, also known as new geocodes, constituted 9.4 percent of the local GIS coordinate updates. Over 98 percent (343,239/348,909) of these address records did not exist on the MTdb prior to the SCEMA program. On the pre-AC file, 85 percent (297,562/348,909) of address records with local GIS coordinate add actions did not have a delivery point type (DPT), thus would not have been DSF adds in time for the 2010 Census AC operation. The ungeocoded address records receiving the local GIS coordinate updates would not have been eligible for the 2010 Census AC operation without the geocodes provided by local government GIS coordinate data. In order for these address records to have been included in the 2010 Census, without the GIS coordinate updates, they would have needed to have been added during the 2010 Census AC operation, or a later operation.

These local GIS coordinate updates originated from local governments, and were therefore concentrated in specific areas, as can be seen in Table A-4 in the Appendix, which displays the local GIS coordinate actions by state. All the addresses with local GIS coordinate updates were observed in twenty-four states, with over half (nearly 52 percent) of those addresses concentrated in only four states.

The largest number of local GIS coordinate updates came from North Carolina, Virginia, Indiana, and Kentucky. While several counties in those four states contained addresses with local GIS coordinates, each of those states contained one county with over 240,000 addresses, for a total of nearly 29 percent of all addresses with local GIS coordinate updates in just those four counties: Mecklenburg, North Carolina; Fairfax, Virginia; Marion, Indiana; and Jefferson, Kentucky. Not surprisingly, these are the largest counties in each state.

There were addresses with local GIS coordinate updates in 81 of the 3,141 counties nationwide. There were ten counties (in nine states) that contributed at least 100,000 addresses each, containing nearly 52 percent of the addresses with local GIS coordinate updates. Five of those ten largest counties were in North Carolina, Virginia, Indiana, or Kentucky.

Table 27 shows that the 2010 Census AC operation verified three-fourths of addresses with local GIS coordinate updates. An additional seven percent of addresses received 2010 Census AC change action codes, which verified that the local GIS-based geocode was correct. According to the 2010 Census AC Assessment, 2010 Census AC verify and change actions made up 68 percent (97,524,708) and 13 percent (18,635,783), respectively, of the 143,356,106 addresses in the eligible 2010 Census AC workload. Addresses with local GIS coordinate updates were more likely to receive a 2010 Census verify action but less likely to receive a 2010 Census change action than the average record eligible for the 2010 Census AC operation.

Table 27. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: GIS Coordinates Address Universe by 2010 Census AC Action Code and Geocode Type

2010 Census AC Action Code	Add Actions (New Geocodes)	Change Actions (Corrected Geocodes)	Total GIS Coordinate-based Actions
Matched to AC add	1,498	6,041	7,539
	(0.4%)	(0.2%)	(0.2%)
Verify	74,265	2,702,442	2,776,707
	(21.3%)	(80.5%)	(74.9%)
Change	27,020	231,970	258,990
	(7.7%)	(6.9%)	(7.0%)
Move	4,033	99,364	103,397
	(1.2%)	(3.0%)	(2.8%)
Delete (verified)	173,374	105,604	278,978
	(49.7%)	(3.1%)	(7.5%)
Single Delete	3	11	14
	(0.0%)	(0.0%)	(0.0%)
Duplicate	30,097	24,424	54,521
	(8.6%)	(0.7%)	(1.5%)
Nonresidential	19,371	30,986	50,357
	(5.6%)	(0.9%)	(1.4%)
Uninhabitable	1,796	6,114	7,910
	(0.5%)	(0.2%)	(0.2%)
No Action	17,452	151,828	169,280
	(5.0%)	(4.5%)	(4.6%)
Total	348,909	3,358,784	3,707,693
	(100.0%)	(100.0%)	(100.0%)

Note: Percentages may not sum to 100 due to rounding.
Source: January 2009 ACS MAFX and the 2010 Census Combo File.

No action codes existed for five percent of the addresses with local GIS coordinate updates. The majority of these addresses were not eligible for the 2010 Census AC filter and thus were not sent to the operation. Three percent of the addresses with local GIS coordinate updates received negative 2010 Census AC actions (duplicate, nonresidential, & uninhabitable) that implied the geocode was correct, compared to 4 percent, or 5,734,825, of the 143,356,106 addresses in the eligible 2010 Census AC workload. Three percent of addresses with local GIS coordinate updates were moves, indicating the geocode was incorrect, compared to 4 percent, or 5,445,699, of the addresses in the eligible 2010 Census AC workload. Deletes removed over seven percent of the addresses with local GIS coordinate updates, compared to 11 percent, or 15,529,825, of

the addresses in the eligible 2010 Census AC workload. Addresses with local GIS coordinate updates were less likely to receive a 2010 Census delete action than the average record eligible for the 2010 Census AC operation.

The majority of the 2010 Census AC outcomes were from the regular AC operation, with only one percent of actions affecting addresses with local GIS coordinate updates coming from the LBAC operation (see Appendix for separate 2010 Census AC and LBAC distributions).

The majority of the addresses with corrected geocodes had positive outcomes, with 87 percent (2,934,412/3,358,784) receiving verify or change actions. Otherwise, those 2.9 million addresses would have been moves during the 2010 Census AC operation, without these corrected geocodes. In comparison, 81 percent (116,160,491) of the 143,356,106 addresses in the eligible 2010 Census AC workload were verify or change actions. So the addresses with corrected geocodes were more likely to receive a positive action (verify or change) than the average record eligible for the 2010 Census AC operation. Three percent (99,364) of the 3,358,784 addresses with corrected geocodes were moved to a different census block during the 2010 Census AC operation. Five percent (167,139/3,358,784) received negative actions (delete, duplicate, nonresidential, or uninhabitable), while another five percent (151,828/3,358,784) were not sent to AC and thus could not be evaluated. Addresses with corrected geocodes were less likely to receive a 2010 Census delete action than the average record eligible for the 2010 Census AC operation.

Half of the addresses with new geocodes (173,377/348,909) were deleted in the 2010 Census AC operation, compared to 11 percent, or 15,529,825, of the 143,356,106 addresses in the eligible 2010 Census AC workload. An additional 16 percent (51,264/348,909) of these address records with local GIS coordinate add actions received other 2010 Census negative actions (duplicate, uninhabitable, or nonresidential) that implied the geocode was correct. Five percent were not sent to the 2010 Census AC operation and thus could not be evaluated. Ungeocoded records are often new construction, and new addresses are often added to the DSF before their corresponding structures are built. As of the 2010 Census AC operation, 85 percent of address records with local GIS coordinate add actions were not yet on the DSF, which may help explain the high delete rate.

Twenty-nine percent (101,285/348,909) of the new geocodes had a positive 2010 Census AC outcome (verify or change actions). When just examining the less than 2 percent (5,670/348,909) of address records with local GIS coordinate add actions with non-SCEMA MAF sources, 86 percent (4,889/5,670) of these records received positive 2010 Census AC actions and 88 percent were in the 2010 Census. These were pre-existing, ungeocoded MTdb records that otherwise would not have been eligible for the 2010 Census AC operation.

Table 28 shows the final 2010 Census validity of the GIS coordinate-based actions. Overall, 85 percent of the 3.7 million addresses with local GIS coordinate updates were in the 2010 Census. Ninety percent of the addresses with corrected geocodes were in the 2010 Census.

Table 28. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: GIS Coordinates Address Universe by 2010 Census Validity and Geocode Type

2010 Census Status	Add Actions (New Geocodes)	Change Actions (Corrected Geocodes)	Total GIS Coordinate-based Actions
In Census	103,150	3,036,794	3,139,944
	(29.6%)	(90.4%)	(84.7%)
Not In Census	245,759	321,990	567,749
	(70.4%)	(9.6%)	(15.3%)
CUF – Invalid	*22,042*	*38,295*	*60,337*
	(6.3%)	*(1.1%)*	*(1.6%)*
CUF – Missing	*223,717*	*283,695*	*507,412*
	(64.1%)	*(8.4%)*	*(13.7%)*
Total	348,909	3,358,784	3,707,693
	(100.0%)	(100.0%)	(100.0%)

Note: Percentages may not sum to 100 due to rounding.
Percentages of Invalid and Missing may not sum to Not in Census due to rounding.
Source: January 2009 ACS MAFX, 2010 Census Combo File, and 2010 Census Unedited File.

The new geocodes were less likely to be included in the 2010 Census, with only thirty percent valid. Of the addresses with local GIS coordinate adds with non-SCEMA MAF sources, 88 percent (4,979/5,670) were in the 2010 Census. Nine percent (31,438/348,909) of address records with local GIS coordinate add actions existed on the pre-AC DSF with DPT values of 'Residential' and 'Included in Delivery Statistics.' Ninety percent (28,328/31,438) of this subset of records with local GIS coordinate add actions were in the 2010 Census, which is consistent with the addresses with corrected geocodes.

The addresses with local GIS coordinate updates were validated in the 2010 Census AC operation at a high rate and 85 percent were valid CUF addresses, so the continued collection of local address files with associated GIS coordinates may prove to be a valuable address list maintenance activity. Updating existing MAF addresses with local GIS coordinates may lead to a smaller number of mis-geocoded units in the MTdb, and thus a higher quality MTdb.

5.4. American Community Survey Time of Interview Actions:

How can we use ACS TOI data to improve address list maintenance?
How frequently were records with ACS TOI delete or other actions confirmed in the 2010 Census AC operation?

The ACS TOI change, delete, and nonresidential actions were recorded on the edited January 2009 ACS MAFX. ACS TOI actions exist in all 50 states, the District of Columbia, and all 3,141 counties. The ACS TOI action codes examined in this report cover all ACS TOI actions recorded on the MTdb up through January 2009, including ACS TOI actions from as far back as Fall 2004.

The ACS filter currently uses ACS TOI delete actions to remove addresses from the ACS sample universe. However, the intelligence of ACS TOI delete actions was not included in the 2010 Census AC filter, and thus, ACS TOI delete actions were included in the 2010 Census AC

operation. The ACS TOI nonresidential actions were not part of either the 2010 Census AC filter or the 2009 ACS filter. As a corollary, ACS TOI actions that verify the address are not sent to GEO to update the MAF and are therefore not identifiable in MAF extracts. The ACS TOI change actions were part of the 2010 Census AC filter, which specified that address records changed by ACS TOI should be marked as valid for the 2010 Census AC universe.

A match between the January 2009 ACS supplemental sample MAFX and the 2010 Census Combo file yielded 390,867 records with ACS TOI change actions, 93,760 records with ACS TOI delete actions and 20,749 with ACS TOI nonresidential actions. Table 29 shows that 90 percent of change actions, 91 percent of delete actions, and less than two-thirds of the nonresidential actions were sent to the 2010 Census AC operation.

Table 29. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: ACS TOI Actions by 2010 Census AC Eligibility

2010 Census AC Filter	ACS TOI Change Actions	ACS TOI Delete Actions	ACS TOI Nonresidential Actions
Eligible	352,956	85,669	13,753
	(90.3%)	(91.4%)	(66.3%)
Not Eligible	37,911	8,091	6,996
	(9.7%)	(8.6%)	(33.7%)
Total	390,867	93,760	20,749
	(100.0%)	(100.0%)	(100.0%)

Note: Percentages may not sum to 100 due to rounding.
Source: January 2009 ACS MAFX and the 2010 Census Combo File.

The majority of the 2010 Census AC outcomes were from the regular 2010 Census AC operation, with less than one-half of one percent of actions resulting from the LBAC operation (see Appendix for separate 2010 Census AC and LBAC distributions).

Table 30 shows that the majority (63 percent) of the ACS TOI change actions received positive AC actions (verify, change, move, match to add). In comparison, 2010 Census AC positive actions (verify, change, move) were 85 percent, or 121,606,190, of the 143,356,106 addresses in the eligible 2010 Census AC workload. Roughly a quarter (28 percent) of the ACS TOI change actions received negative AC actions (delete, duplicate, nonresidential, uninhabitable), while the 2010 Census AC negative actions were 15 percent, or 21,264,549, of the 143,356,106 addresses in the eligible 2010 Census AC workload.

Records with ACS TOI change actions were removed during the 2010 Census AC operation at a higher rate than other records in the 2010 Census AC workload, though a majority of the records with ACS TOI change actions were confirmed as good records during the 2010 Census AC operation. Recall that the study universe includes all ACS TOI change actions since 2004. Conditions on the ground might have changed since ACS FRs visited these addresses. Resource constraints did not permit longitudinal analysis to estimate the age of the ACS TOI action from when each action first appeared on the MTdb.

Table 30. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: ACS TOI Actions by 2010 Census AC Action Code

AC results	ACS TOI Change actions	ACS TOI Delete Actions	ACS TOI Nonresidential Actions
Matched to AC add..	1,820	96	571
	(0.5%)	(0.1%)	(2.8%)
Verify	124,547	15,019	3,060
	(31.9%)	(16.0%)	(14.7%)
Change.....	114,609	9,958	2,567
	(29.3%)	(10.6%)	(12.4%)
Move.....	4,835	823	98
	(1.2%)	(0.9%)	(0.5%)
Delete (verified).....	63,534	52,509	3,940
	(16.3%)	(56.0%)	(19.0%)
Single Delete	2	0	0
	(0.0%)	(0.0%)	(0.0%)
Duplicate	39,580	4,801	1,018
	(10.1%)	(5.1%)	(4.9%)
Nonresidential	1,979	1,358	2,676
	(0.5%)	(1.4%)	(12.9%)
Uninhabitable	2,913	989	275
	(0.7%)	(1.1%)	(1.3%)
No action	37,048	8,207	6,544
	(9.5%)	(8.8%)	(31.5%)
Total records.....	390,867	93,760	20,749
	(100.0%)	(100.0%)	(100.0%)

Note: Percentages may not sum to 100 due to rounding.
Source: January 2009 ACS MAFX and the 2010 Census Combo File.

Table 30 shows that over half of the ACS TOI deletes were also deleted during the 2010 Census AC operation. Positive 2010 Census AC actions (verify, change, move, matched to add) were assigned to 28 percent of ACS TOI deletes, while the remaining 72 percent received negative 2010 Census AC actions (delete, duplicate, nonresidential, uninhabitable) or no 2010 Census AC action. In comparison, 2010 Census AC deletes were 11 percent, or 15,529,724, of the 143,356,106 addresses in the eligible 2010 Census AC workload and all 2010 Census AC negative actions (delete, duplicate, nonresidential, uninhabitable) were 15 percent, or 21,264,549, of the addresses in the eligible 2010 Census AC workload. ACS TOI deletes were deleted at a higher rate than other records going into the 2010 Census AC operation and a majority of these records had negative outcomes.

Of the roughly two-thirds of the ACS TOI nonresidential actions that either went to the 2010 Census AC operation or were matched to 2010 Census AC adds, fewer (30 percent of all nonresidential actions) received positive AC actions (verify, change, move, matched to add).

For Table 31, the ACS TOI actions were matched to the CUF to determine their final 2010 Census validity. Sixty-two percent of the ACS TOI change actions were in the 2010 Census. Meanwhile, the negative ACS TOI actions were less likely to be in the 2010 Census, with 23 percent of the ACS TOI delete actions and 24 percent of the ACS TOI nonresidential actions valid in the 2010 Census.

Table 31. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: ACS TOI Actions by 2010 Census Validity

2010 Census Status	ACS TOI Change Actions	ACS TOI Delete Actions	ACS TOI Nonresidential Actions
In Census	242,197	21,739	4,893
	(62%)	(23%)	(24%)
Not In Census	148,670	72,021	15,856
	(38%)	(77%)	(76%)
CUF – Invalid	*11,034*	*7,237*	*1,793*
	(3%)	*(8%)*	*(9%)*
CUF – Missing	*137,636*	*64,784*	*14,063*
	(35%)	*(69%)*	*(68%)*
Total	390,867	93,760	20,749
	(100%)	(100%)	(100%)

Note: Percentages may not sum to 100 due to rounding.
Source: January 2009 ACS MAFX, 2010 Census Combo File, and 2010 Census Unedited File.

In answering the research question, the ACS TOI actions may be useful for address list maintenance, and should be considered in future address filters, targeting methods, or data mining methods. While not an overwhelming association, the delete and nonresidential actions correlate with invalid records, and the change actions are associated with valid records.

6. Related Evaluations and Assessments

The following 2010 Census Program for Evaluations and Experiments reports and 2010 Census Assessments are related to the research conducted in this evaluation:

- Evaluation of Address Frame Accuracy and Quality (Johnson and Kephart, 2013)
- Study of Address Canvassing Targeting and Cost Reduction (Boies, Shaw, and Holland, 2012)
- Evaluation of Data-Based Extraction Processes for the Address Frame (Ward, 2012)
- Evaluation of Small Multi-Unit Structures (Virgile, 2012)
- 2010 Census Address Canvassing Operational Assessment (Address List Operations Implementation Team, 2012)

7. Conclusions and Recommendations

7.1. Conclusions

7.1.1. Geocode Imputations

The geocode imputations obtained from commercial software for this evaluation have utility for numerous purposes, and for numerous MTdb stakeholders. It is clear there is a strong relationship between the type of geocode imputation and the accuracy of the geocode. Point geocode imputations and best address geocode imputations provide highly reliable census geocodes.

When the universe of 6,440,043 ungeocoded MTdb address records was studied (a universe selected on the condition that had a specific address record contained a geocode, it would have been valid for the 2010 Census AC operation or the 2009 ACS sample) for its overlap with the 2010 Census AC operation:

- 97 percent (1,033,090/1,064,701) of Point geocode imputations were found to be accurate,
- 81 percent (760,164/937,615) of Best Address geocode imputations were found to be accurate and
- 56 percent (139,170/247,790) of Other Address, Best ZIP Code geocode imputations were found to be accurate.

Further, to remove the bias associated with the 2010 Census AC operation eligibility and outcome requirements, a supplemental field operation was designed and conducted. The results of this supplemental field operation indeed confirmed the reliability of the Point and Best Address geocode imputations. Point and Best Address geocode imputations represent 62.1 percent of the study universe (3,998,531 of the 6,440,043 previously ungeocoded, and potentially valid, address records nine years after the last census) and would have represented a 2.8 percent (3,998,531/143,356,106) increase in the eligible 2010 Census AC workload if the imputed geocodes had been applied to the address list. The results of the nationally-representative field-validated sample of 6,549 address records showed:

- 90.6 percent (89.2, 92.1) of Point geocode imputations were found to be accurate, and
- 73.4 percent (71.9, 74.8) of Best Address geocode imputations were found to be accurate,

for a 90 percent confidence interval (the lower and upper bound follow the percentages). Lastly, address records with other geocode imputation types (other ZIP Code imputations) were not evaluated further, primarily due to their by-design low level of precision and low concordance with the 2010 Census AC operation geocoding results.

We expect that ungeocoded but potentially valid records will gradually increase approaching the next census, and a reliable method for geocoding a majority of these records is needed. Geostan may be a good complement to our internal geocoding activities. Note that the most reliable Geostan geocodes were derived from coordinate data, such as collected in the geocodes derived from local GIS address coordinates. Active participation by local governments in an ongoing collection of local GIS address coordinates may overlap the potential gain from commercial software in the coming decade.

7.1.2. Administrative Records

AR data from StARS 2007 and StARS 2008 were researched to determine their utility for contributing to MAF maintenance activities. All 184,300,099 MTdb address records with non-blank address types and ZIP codes of length five available in the 2010 Census pre-AC and post-AC universes were selected for analysis, including all unique location and mailing addresses. Once expanded, up to four unique addresses per MAFID (to maximize the matching outcomes), 224,927,714 unique MTdb addresses, were matched at the unit-level to the StARS 2007 and StARS 2008 databases and unduplicated. The StARS 2007 and StARS 2008 MHFs contained approximately 152 million and 154 million unique address records respectively.

Overall, of the 224,927,714 unique address records in the MTdb studied here, 119,427,271 or 53 percent matched to StARS. Of the 152,203,709 unique 2007 StARS addresses, 112,509,044 matched to the MTdb, and of the 154,343,213 unique 2008 StARS addresses, 113,960,259 matched to the MTdb. Almost 100 percent of the address matches were city-style addresses.

In order to assess whether or not AR data provide intelligence for adding new address records to the MTdb, the DSSD examined how many of the 2010 Census AC operation adds were contained in StARS 2007 and StARS 2008. Of the approximately 9.9 million adds in the 2010 Census AC operation, StARS contained 67 percent (2,741,141/4,071,530) of the adds that matched back to the MAF and five percent (265,163/5,836,247) of the new adds. This confirms that AR data can provide independent evidence of address validity for the universe of address records excluded from the 2010 Census AC filter. If valid AR addresses could be identified in an efficient and accurate manner prior to the listing operation, this AR data potential would have represented a possible addition of over two percent (3,006,304) of the 143,356,106 addresses in the eligible 2010 Census AC operation workload.

Identification of potential adds that do not meet the current filter rules are not in the MTdb could focus efforts to improve the MTdb on records, or areas containing records, that might be targeted for field operations or other address list maintenance activities. For instance, if these potentially valid records are ungeocoded, they could be processed through imputation software (see previous research topic) to see if sufficiently high quality imputed geocodes are available. Alternatively, the Census Bureau could request governments with significant numbers of these records to consider submitting address files with GIS coordinates (see next research topic).

In order to assess whether or not AR data provide intelligence for confirming existing address frame records, the DSSD examined how many of the records with positive 2010 Census AC operation action codes were contained in StARS 2007 and StARS 2008. StARS contained: 88 percent (85,909,962/97,511,328) of all 2010 Census AC verify address records; 88 percent (4,813,958/5,445,667) of all 2010 Census AC move address records; and 69 percent (12,639,561/18,444,809) of all 2010 Census AC change address records. This confirms that AR data can provide independent evidence of address validity for the universe of address records included from the 2010 Census AC filter. The 2010 Census AC operation cost nearly 845 million dollars, including field and contract overhead costs (Holland, 2012), and therefore offers high cost reduction potential. This AR data potential, if valid AR addresses could be identified in an efficient and accurate manner prior to the listing operation, would have represented a possible reduction of over 72 percent (103,363,481) to the 143,356,106 addresses in the eligible 2010 Census AC operation workload.

Additionally, assessing whether or not AR data provide intelligence for identifying addresses that should be removed from the address frame, the DSSD examined how many of the records with negative 2010 Census AC operation action codes were contained in StARS 2007 and StARS 2008. Of all address records that received negative action codes in the 2010 Census AC (delete, duplicate, nonresidential, uninhabitable), 73 percent (14,521,685/19,889,845) were absent from both StARS 2007 and StARS 2008. In particular, 74 percent (10,589,737/14,692,445) of address records with 2010 Census AC delete actions were not present in both StARS 2007 and StARS 2008. This AR data potential, had an efficient and accurate manner of identifying invalid MTdb addresses via AR been developed prior to the listing operation, would have represented a possible reduction of over 10 percent (14,521,685) to the 143,356,106 addresses in the eligible 2010 Census AC operation workload.

The presence of an address in StARS increases the likelihood that the address was found to be valid in the 2010 Census AC operation. Of the approximately 151 million addresses in the 2010 Census AC workload[35], about 131 million addresses, or 87 percent, were valid. For the approximately 112 million addresses that matched StARS, about 106 million addresses, or 95 percent, were valid in the 2010 Census AC operation. For this universe of address records, a match to the StARS database increases the likelihood of a positive listing outcome. In a similar manner, the absence of an address record in StARS increases the likelihood of a negative 2010 Census AC outcome (delete, nonresidential, uninhabitable). Of the 151 million addresses in the 2010 Census AC workload, there were about 20 million addresses (13 percent) with negative listing outcomes. For the 39 million addresses that did not match StARS, about 15 million addresses, or 37 percent, received negative listing outcomes in the 2010 Census AC operation. For this universe of address records, the absence of a match to the StARS database increased the likelihood of a negative listing outcome by 2.7 times that observed in the overall operation workload.

The 2010 Census AC filter offered an 83.5 percent (153,866,943/184,300,099) correct prediction of the 2010 CUF, while the StARS match status offered an 80.7 percent (148,705,784/184,300,099) correct prediction of the 2010 CUF. This net difference of 2.8 percent (5,161,159/184,300,099) in predicting the correct final 2010 Census status indicates a strong similarity in predicting the final 2010 Census status between the StARS match status and 2010 Census AC filter.

The 2010 Census AC filter included LUCA records in order to offer feedback to participating governments, but resources may not permit such a large field operation for future decennial censuses, so an alternative method of feedback may be needed. Since the StARS match status and the 2010 Census AC filter had similar outcomes through the 2010 Census, the StARS match status offers a potential alternative to a national field operation.

[35] The 2010 Census AC workload is all 151,199,426 address records that were eligible for StARS matching and that received 2010 Census AC operation actions (including all add actions).

In summary, the strong relationships demonstrated here are very useful – where the presence of an address in StARS increases the likelihood of a positive listing outcome and positive final census status, and its obversion[36]: the absence of an address in StARS increases the likelihood of a negative listing outcome and negative final census status. From these findings, StARS and other AR data sources will undoubtedly prove very useful in future data modeling, data mining, microsimulations, and targeting research to reduce costs and maintain accuracy during future censuses and surveys.

7.1.3. Local GIS coordinates

In assessing the accuracy of geocodes derived from local GIS address coordinates, the DSSD studied a universe of approximately 3.7 million addresses. This universe consists of MTdb records flagged with SCEMA action codes derived from local address files with associated GIS coordinate data received from MTAIP. The GEO used these local GIS coordinates to identify geocodes for ungeocoded addresses (new geocodes) and to verify or correct geocodes on existing geocoded addresses in the MTdb. This quantity of address records represents 2.6 percent (3,707,693/143,356,106) of the eligible 2010 Census AC workload.

The majority of the addresses with corrected geocodes sent to the 2010 Census AC operation had a positive outcome, with 87 percent receiving verify or change actions. These 2.9 million addresses would have been moves during the 2010 Census AC operation without their corrected geocodes. Recall that a move is a post-processing outcome requiring both a 2010 Census AC add and a 2010 Census AC delete action in the field operation. Addresses with corrected geocodes were more frequently validated (81 percent, or 2,702,442/3,358,784) than addresses with new geocodes (21 percent, or 74,265/348,909).

Half of the addresses with new geocodes (173,374/348,909) were deleted in the 2010 Census AC operation, a much higher delete rate than for all 143,356,106 addresses eligible for the 2010 Census AC operation at 10 percent (or 15,529,724). Ungeocoded records are often new construction, and addresses are often added to the DSF before their corresponding structures are built. Similarly, local governments often keep track of new residential addresses at the permit stage (before construction begins). As of the 2010 Census AC operation, 85 percent of address records with local GIS coordinate add actions were not yet on the DSF, which may help explain the high delete rate.

Knowledge of a building's existence, for an address with local GIS coordinates that does not exist on the MTdb, would allow storage of the new address's geocode, while filtering it out of surveys and the decennial census, so it does not increase field costs. Some local governments offer building footprint files online, so collecting information from local governments indicating whether an address record reflects an existing building or a building under construction (and a

[36] Obversion is a logic term referring to a form of inference whereby a negative statement is derived from a positive statement or vice versa. An example is "None of these addresses are ungeocoded" from "All of these addresses are geocoded."

year expected to finish construction if known) may be possible. Alternately, existence on the DSF, alongside delivery point type values of 'Residential' and 'Included in Delivery Statistics', could indicate building existence for address filters.

The 2010 CUF provided results similar to the 2010 Census AC action codes. Ninety percent (3,036,794/3,358,784) of the addresses with corrected geocodes were in the 2010 Census, while only thirty percent (103,150/348,909) of the addresses with new geocodes were included in the 2010 Census. Updating existing MAF addresses with local GIS coordinates may lead to a smaller number of mis-geocoded units in the MTdb, and thus a higher quality MTdb.

7.1.4. American Community Survey Time of Interview Actions

The purpose of this research topic was to investigate whether ACS TOI actions should be included in address filters. The ACS filter currently uses ACS TOI delete actions to remove addresses from the ACS sample universe. However, the intelligence of ACS TOI delete actions was not included in the 2010 Census AC filter, and thus, addresses with ACS TOI delete actions were included in the 2010 Census AC operation. The ACS TOI change, delete, and nonresidential actions were recorded on the edited January 2009 ACS MAFX. There are 390,867 records with ACS TOI change actions, 93,760 records with ACS TOI delete actions and 20,749 records with ACS TOI nonresidential actions, amounting to 0.27 percent, 0.07 percent, and 0.01 percent of the 2010 Census AC eligible workload respectively.

The majority (63 percent, or 245,811/390,867) of the addresses with ACS TOI change actions received positive AC actions (verify, change, move, matched to add), and a similar percentage (62 percent, or 242,197/390,867) of records were in the 2010 Census. Over half of the addresses with ACS TOI deletes (56 percent, or 52,509/93,760) were also deleted during the 2010 Census AC operation, for a total of 72 percent (67,864/93,760) that received negative 2010 Census AC actions (delete, duplicate, nonresidential, uninhabitable) or no 2010 Census AC action. Similarly, 77 percent, or 72,021/93,760, were not in the 2010 Census. The addresses with ACS TOI nonresidential action results were similar overall to the ACS TOI deletes, with 70 percent (14,453/20,749) negative 2010 Census AC actions (delete, duplicate, nonresidential, uninhabitable) or no 2010 Census AC action, and 76 percent (15,856/20,749) that were not in the 2010 Census.

7.2. Recommendations

This evaluation found that using information internally available, but not currently incorporated into the updating process for the MTdb, would assist the Census Bureau in improving its address frame for future censuses and surveys. Based on these findings, the DSSD presents the following recommendations:

1) **Conduct a Cost Benefit Analysis for Geocoding Activities.** We recommend the GEO conduct a cost benefit analysis to assess the return on investment for licensing, renewing, and maintaining trained support staff to operate third-party geocoding software. The software is a substantial investment each year it is licensed, and foregoing license renewal for years when the software is unnecessary may incur additional charges; however, these costs are less than previous decennial address listing/canvassing activities. The GEO should quantify the quantities of address records geocoded by each of the internal geocoding activities, as well as development and maintenance costs of each (assuming internal geocoding activities continue). This formal documentation should quantify the error rates of external and internal geocoding activities and the tradeoffs of not conducting any geocode imputations or accepting a measured level of error in geocoding activities. Measuring the size and growth of the ungeocoded universe biannually (with the delivery sequence file updates) is essential. This biannual cost benefit analysis is necessary to inform budgeting and planning activities for the next Census.

2) **Apply Point and Best Address Geocode Imputations to the MTdb as Warranted by the Cost Benefit Analysis.** If the previous cost benefit analysis warrants purchase or renewal of the software, we recommend processing all ungeocoded address list records through the third-party software to capture only the Point and Best Address imputations. These geocodes need to be recalculated from the imputed latitude and longitude coordinates using GEO shape files. These final imputed, recalculated geocodes are reliable and accurate for MTdb updates. Both the geocode imputation and geocode imputation type (accuracy level) should be housed in the MTdb. We recommend that MTdb stakeholders be provided with documentation on the third-party geocode imputation software, and consider the usefulness of Other Address imputations and the Best ZIP Code imputations in specific applications. At this time, given the low level of accuracy, we do *not* recommend the use of Other ZIP Code imputations for geocoding ungeocoded records. This process should be repeated, as warranted by the biannual cost benefit analysis, for all ungeocoded records in the MTdb at that time. This universe will be largely comprised of new DSF adds and newly included in delivery statistics records.

3) **Periodically Evaluate the Accuracy of the MTdb Geocode Imputations.** Since MTdb maintenance procedures will change over time, and future MTdb input data sources will vary in both quantity and quality, we recommend periodic evaluation of any geocode imputations applied to the MTdb at intervals not to exceed two years. This evaluation work should be consistent with the analyses conducted in this evaluation, examining both the level of concordance with previous field activities and operations, and the level of accuracy measured by a supplemental field operation that investigates the universe of records for which the Census Bureau does not currently possess a geocode.

4) **Include Address List Geocode Imputations in Address Update Targeting Research.** The geocode imputations research results should be incorporated into the targeted address canvassing research. Geographical areas with highly accurate geocode imputations can be updated using imputation software rather than fieldwork. However, geographical areas with a majority of inaccurate imputations and numerous ungeocoded addresses are strong candidates for fieldwork targeted to locate the ungeocoded addresses and improve address coverage for those areas.

5) **Use Locate Address Assignments for Research and other Field Activities, Including Address Update Targeting Research.** The LAA worked well for our geocode imputation research, and are a useful method to collect data for field activities, research, and targeting focused on distinct addresses or groups of addresses, rather than a complete block listing. If only a few addresses in a block are in need of updating or of interest for research, LAA is a more efficient way of collecting information than listing every address in a block. The new corporate listing device should include this type of assignment capability and utilize navigational routing.

6) **Construct a Single Integrated Administrative Record Data Repository.** The process of acquiring, building, integrating, and conducting analysis on the large data files for the AR research topic required a substantial effort. If the matching and integration activities are done once at an agency level, similar to the integration and maintenance of MTdb and the Business Register, all research and production activities would benefit significantly. Further, this integrated AR data repository should include both decennial data and regular MTdb updates so that users do not need to allocate significant amounts of time to merging this integrated AR data repository with the latest MTdb extracts. Identifying a full-time team of statisticians and IT specialists to create and maintain a single integrated AR data repository which combines AR data with recent and historical MTdb extracts and decennial data will streamline this process.

7) **Expand Access to Administrative Records for Research.** All AR research and data acquired and integrated per the last recommendation should be made as widely available to the rest of the Census Bureau as possible. The results in this evaluation show that the presence of addresses in StARS increases the likelihood for positive listing actions, and the absence of addresses in StARS increases the likelihood for negative listing actions. Given these strong relationships, the DSSD recommends providing StARS and other administrative records data accessibility to all areas in the Census Bureau conducting decennial census and current survey address list coverage and improvement research. All areas in the Census Bureau should have access to StARS and other administrative records data for data modeling, data mining, and microsimulation research. Divisions involved in decennial census or survey operations should employ Title 26 trained personnel who are both authorized for and involved in active AR research with direct access to AR data in Title 26 secured areas. Having trained personnel with direct experience with the interactions between AR data and the decennial census or survey work of their division will optimize the development of specific methods to use AR data in census and survey operations. We recommend the pursuit of inter- and intra-agency approval for the

following expansion of decennial census research using StARS and other AR data: address frame filter research, targeted address canvassing research, improving Listing Quality Control methods, and general decennial census cost reduction research (for Address Canvassing and Nonresponse Followup).

8) **Update the MTdb with Administrative Records Flags.** As measured in this evaluation, data from StARS enhance the Census Bureau's ability to identify valid addresses, and are therefore useful for address frame updating and maintenance. Flags indicating whether a StARS match existed for each MTdb record should be included on the MTdb, with annual updates of matches to StARS and other AR data, similar to the existing DSF flags. Documentation detailing the meaning of each administrative records flag should be widely available for collaborative, cross-divisional research efforts by technical and subject experts to speed the development and implementation of practical administrative records use. These flags would be an ongoing method to incorporate AR data into decennial census and current survey work. The presence of a series of StARS match flags would allow MTdb users to include the presence or absence of StARS matches in their filter for specific applications, such as data modeling, data mining, and microsimulation work.

9) **Collect Local GIS Coordinates.** As measured in this evaluation, quality, local GIS geocode corrections were highly reliable. Continued collection of these local GIS coordinates to verify or correct geocodes on existing geocoded addresses may lead to a smaller number of mis-geocoded units in the MTdb. This translates into a higher quality MTdb and reduced field costs due to mis-geocoded addresses. The costs of this continued collection program should be quantified for previous activities, and estimated for future ones to calculate the return on investment for this program.

10) **Collect Local Building Information along with Local GIS Coordinates.** Half of the addresses with new geocodes derived from local GIS coordinates were deleted in the 2010 Census AC operation. A 'Provisional Add' unit status for an address with local coordinates that was not on the address list would allow storage of the new address's geocode, while filtering it out of surveys and the decennial census, so it does not increase field costs. Some local governments offer building footprint files online, so collecting information from local governments indicating whether an address record reflects an existing building or a building under construction (and a year expected to finish construction if known) may be possible and whether this information offers a good return on investment should be investigated. Alternately, existence on the DSF, alongside a DPT of 'Residential' and 'Included in Delivery Statistics,' could be used as a filter indicator for building existence. In an ongoing collection of addresses with local GIS coordinates, if local governments would include updated information on whether a building exists for a particular address record in a timely manner, then the MTdb would better reflect actual ground conditions.

11) **Further Investigate ACS TOI Actions.** The ACS TOI actions are a small portion of the MTdb but may still be useful for address list maintenance; and the results are not decisive enough to require their inclusion in address filters at this time. Further research should be

considered to improve understanding of these action codes, their characteristics, and their associated validity.

8. Acknowledgements

Many people contributed to the successful completion of this report. My branch (John Boies, Jonathan Holland, Nancy Johnson, Kathleen Kephart, Kevin Shaw, Matthew Virgile, and Justin Ward) put in many long hours discussing methodology, SAS® coding & troubleshooting, analysis, and offered abundant feedback on the statistical analysis and supporting prose. I am quite grateful to have such patient, thoughtful, and intelligent co-workers. Thanks to Jennifer Reichert for support, facilitation, feedback, and paring down the research topics to a manageable level. Thanks to Chuck Whittington, Hongyu Doan, Nicholas Doner, and Dan Keefe for their help in providing geocode imputations for the potentially valid ungeocoded records. Thanks to Deb Wagner, Chris Boniface, Diana Liu, Aneesah Williams, Teresa Schellhamer and others in DID for their work matching and delivering data from StARS. Thanks to Jacqueline Jones, John Godenick, Leo Feighny, Nathan Hillila, and Feder Estelus for organizing the field work and communicating with the RO staff. Thanks to Gina Winchester and staff for providing LAA field training materials and providing feedback in editing materials. Thanks to the RO staff for organizing and carrying out the field work. Thanks to Jim Hartman, Alan Peterson, Holly Malanowski, and Stephen Clawson for obtaining OMB clearance for the field work and for delivering data to and from TMO. And many, many thanks to Claude Jackson for his endless and invaluable IT support.

Title 26 IRS data was used for the AR research topic in this evaluation. Throughout the process to identify, build-out, and secure separate office space to conduct this research using Title 26 data, we owe many, many thanks to Kevin Shaw, Jennifer Reichert, David Whitford, Dan Weinberg, Tom Mesenbourg, Susan Boyer, John Fisher, Bob Drew, and others. Additionally, it was necessary to stand up a secure server to conduct this research. For this activity, we owe many thanks to Claude Jackson, Kevin Shaw, Jennifer Reichert, David Whitford, James Trigg, Curtis Broadway, Pam Mosley, Arnold Jackson, and numerous others in the Decennial Systems and Contract Management Office and the Information Technology Security Office. Without the help of many dedicated employees to help establish a secure office and server, this research would not have occurred.

9. References

Address List Operations Implementation Team (2012), "2010 Census Address Canvassing Operational Assessment," 2010 Census Planning Memoranda Series Number 168, January 10, 2012.

Bates, Lawrence (2009), "Editing the MAF Extracts and Creating the Unit Frame Universe for the American Community Survey (2009 Main Phase)," DSSD American Community Survey Memorandum Series Chapter Universe Creation ACS09-UC-1, January 5, 2009.

Boies, John L. (2012), Kevin M. Shaw, and Jonathan P Holland, "2010 Census Program for Evaluations and Experiments (CPEX): Study of AC Targeting and Cost Reduction, Report," DSSD 2010 Census Program for Evaluations and Experiments Memorandum Series A-09, July 12, 2012.

Bureau of the Census (2009a), "American Community Survey Design and Methodology," Washington, D.C., April 2009.

Bureau of the Census (2009b), "Introduction to the 2020 Census," Washington, D.C., June 18, 2009.

Bureau of the Census (2010), "Strategic Plan for the 2020 Census," Washington, D.C., Version 2.0, September 30, 2010.

Clark, Sonja, (2009), "2010 Census Program for Evaluations and Experiments Study Plan: Evaluation of Data-Based Extraction Processes for the Address Frame," DSSD 2010 Decennial Census Memorandum Series O-A-6, August 13, 2009.

Dixon, Kelly (2007), and Christine Tomaszewski, "Quality Control Sampling Procedures for the Structure Coordinate Enhancement of the MAF (Master Address File) Addresses (SCEMA) Project Memorandum," Census Bureau, February 22, 2007.

Garcia, Mayra (2009), "2010 Census Program for Evaluations and Experiments Study Plan: Study of Address Canvassing Targeting and Cost Reduction," DSSD 2010 Decennial Census Memorandum Series G-12, September 29, 2009.

Gordon, Judith J. (2008), "Census 2010: Dress Rehearsal of Address Canvassing Revealed Persistent Deficiencies in Approach to Updating the Master Address File," Final Report No. OSE-18599, Office of Inspector General, October 2008.

Hakanson, Amanda S. (2008), Nancy R. Johnson, Christine M. Tomaszewski and Sonja K. Clark, "Address Coverage Improvement and Evaluation Program – 2007 National Estimate of Coverage of the Master Address File," September 30, 2008.

Holland, Jonathan (2009), "Evaluation of GeoStan on Ungeocoded Records from the 2008 Census Dress Rehearsal," Census Bureau, draft last updated in April 28, 2009.

Holland, Jonathan, P. (2009), and Matthew Virgile, "2010 Census Program for Evaluations and Experiments Study Plan: Study of Automation in Field Data Collection for Address Canvassing," DSSD 2010 Decennial Census Memorandum Series O-A-2, November 24, 2009.

Holland, Jonathan P (2012), "2010 Census Program for Evaluations and Experiments (CPEX): Evaluation of Automation in Field Data Collection in Address Canvassing Report," DSSD 2010 Census Program for Evaluations and Experiments Memorandum Series A-05, July 24, 2012.

Johnson, Nancy, (2011), "2010 Census Program for Evaluations and Experiments Study Plan: Evaluation of Address Frame Accuracy and Quality," DSSD 2010 Decennial Census Memorandum Series O-A-3R, June 2, 2011.

Johnson, Nancy, (2013) and Kathleen Kephart, "2010 Census Program for Evaluations and Experiments: Evaluation of Address Frame Accuracy and Quality Report," DSSD 2010 Census Program for Evaluations and Experiments Memorandum Series A-07, 2013.

Kwiat, Aliza, (2009), "Examining Blocks with Lister Error in Area Listing", pages 2546-2557 of Proceedings of the Section on Survey Research Methods, American Statistical Association, August 4, 2009.

Owens, Karen (2009a), "Customer Requirements for the 2010 Address Canvassing Products Delivery," DSSD 2010 Decennial Census Memoranda Series, No. D-01R3, U.S. Census Bureau, January 15, 2009.

Owens, Karen (2009b), "Customer Requirements for the Address Filter for the 2010 Census Initial UC&M and GQV Address Products," DSSD 2010 Decennial Census Memoranda Series, No. D-07, U.S. Census Bureau, February 13, 2009.

Perrone, Susan (2005), "Address Coverage Improvement and Evaluation Program – Assessment of the Demographic Area Address Listing Program," Census Bureau, October 24, 2005.

Ruhnke, Megan (2003), "Census 2000 Evaluation F.15: An Assessment of the Addresses on the Master Address File Missing in the Census or Geocoded to the Wrong Collection Block," Census Bureau, August 19, 2003.

Tomaszewski, Christine (2009), "Customer Requirements for Geocoding of Ungeocoded Records for the Evaluation of Address List Maintenance Using Supplemental Data Sources," DSSD 2010 Decennial Census Memoranda Series, No. K-01, U.S. Census Bureau, June 19, 2009.

Tomaszewski, Christine Gibson (2010a), "Customer Requirements Documentation for StARS Results to support the Evaluation of Address List Maintenance Using Supplemental Data Sources," DSSD 2010 Decennial Census Memorandum Series O-A-1, March 22, 2010.

Tomaszewski, Christine Gibson (2010b), "2010 Census Evaluation Study Plan: Evaluation of Address List Maintenance Using Supplemental Data Sources," DSSD 2010 Decennial Census Memorandum Series O-A-4, December 14, 2010.

Virgile, Matt, (2010), "2010 Census Program for Evaluations and Experiments Study Plan: Evaluation of Small Multi-Unit Structures," DSSD 2010 Decennial Census Memorandum Series O-A-14, January 11, 2010.

Virgile, Matt (2012), "2010 Census Program for Evaluations and Experiments: Evaluation of Small Multi-Unit Structures Report," DSSD 2010 Census Program for Evaluations and Experiments Memorandum Series A-01, February 21, 2012.

Ward, Justin (2011), "Documentation for the 2010 Census Address Frame Combination File, Version #1," DSSD 2010 Decennial Census Memorandum Series D-19, August 24, 2011.

Ward, Justin (2012), "2010 Census Program for Evaluations and Experiments (CPEX) Evaluation of Data-Based Extraction Processes for the Address Frame," DSSD 2010 Census Program for Evaluations and Experiments Memorandum Series A-04, June 27, 2012.

Appendix

Table A-1. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: Caseload and Timing of Fieldwork by Regional Office

Regional Office	Phase II Beginning Date	Phase II Sample Cases	All Sample Cases
Boston	8/16/2011	279	348
New York	8/12/2011	30	84
Philadelphia	8/16/2011	344	444
Detroit	8/14/2011	311	370
Chicago	8/23/2011	269	344
Kansas City	8/12/2011	651	717
Seattle	8/12/2011	317	417
Charlotte	8/12/2011	707	793
Atlanta	8/18/2011	857	957
Dallas	8/19/2011	1,302	1,402
Denver	8/15/2011	370	470
Los Angeles	8/8/2011	103	203
Total	8/8/2011	5,540	6,549

Source: Emails, Pre-AC file, Post-AC file, January 2009 ACS Extract, and GeoStan Imputations File.

Table A-2. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: Caseload and Results of Fieldwork by Regional Office

Regional Office	Results File	Missing Cases	All Sample Cases	Percent (of All Sample Cases)
Atlanta	953	4	957	14.6
Boston	347	1	348	5.3
Charlotte	751	42	793	12.1
Chicago	344	0	344	5.3
Dallas	1,376	26	1,402	21.4
Denver	462	8	470	7.2
Detroit	364	6	370	5.6
Kansas City	712	5	717	10.9
Los Angeles	188	15	203	3.1
New York	84	0	84	1.3
Philadelphia	438	6	444	6.8
Seattle	413	4	417	6.4
Total	6,432	117	6,549	100.0

Source: Emails, Pre-AC file, Post-AC file, January 2009 ACS Extract, and GeoStan Imputations File.
Note: Percentages may not sum to 100 due to rounding.

Table A-3. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: Potentially Valid Ungeocoded Records with Imputed Blocks by State for Records Found before start of GQV operation

State	Imputed Geocode and 2010 Census AC Geocode Agreement (%Total Ungeocoded Addresses in 2010 Census AC)	Total Ungeocoded Addresses in 2010 Census AC (% of Total Geocode Imputations Universe)	Total Geocode Imputations Universe (% of United States)
Alabama	47,481 (57.5%)	82,564 (37.3%)	221,299 (3.4%)
Alaska	3,836 (68.2%)	5,627 (51.8%)	10,853 (0.2%)
Arizona	27,409 (76.8%)	35,699 (33.4%)	106,770 (1.7%)
Arkansas	19,462 (53.1%)	36,686 (40.5%)	90,648 (1.4%)
California	162,917 (87.3%)	186,544 (55.9%)	333,904 (5.2%)
Colorado	22,227 (81.8%)	27,176 (32.8%)	82,822 (1.3%)
Connecticut	8,018 (89.3%)	8,974 (34.3%)	26,148 (0.4%)
Delaware	28,354 (91.6%)	30,948 (39.6%)	78,153 (1.2%)
District of Columbia	342 (85.7%)	399 (33.7%)	1,183 (0.0%)
Florida	132,290 (79.1%)	167,259 (42.5%)	394,011 (6.1%)
Georgia	104,758 (66.3%)	157,953 (41.3%)	382,661 (5.9%)
Hawaii	11,280 (85.5%)	13,189 (63.1%)	20,887 (0.3%)
Idaho	14,764 (63.3%)	23,329 (47.4%)	49,167 (0.8%)
Illinois	33,071 (68.0%)	48,613 (37.2%)	130,616 (2.0%)
Indiana	34,681 (61.3%)	56,551 (45.5%)	124,375 (1.9%)
Iowa	12,126 (62.3%)	19,478 (35.2%)	55,314 (0.9%)
Kansas	11,357 (59.0%)	19,247 (35.0%)	54,989 (0.9%)
Kentucky	25,231 (39.7%)	63,475 (40.6%)	156,463 (2.4%)
Louisiana	23,944 (54.2%)	44,172 (46.2%)	95,606 (1.5%)
Maine	24,119 (65.3%)	36,934 (45.4%)	81,401 (1.3%)
Maryland	11,900 (65.3%)	18,237 (40.1%)	45,524 (0.7%)
Massachusetts	13,736 (75.3%)	18,251 (41.2%)	44,336 (0.7%)
Michigan	40,405 (69.4%)	58,211 (40.0%)	145,514 (2.3%)
Minnesota	35,556 (69.4%)	51,262 (46.9%)	109,274 (1.7%)
Mississippi	26,486 (44.2%)	59,962 (41.6%)	144,201 (2.2%)
Missouri	42,555 (56.8%)	74,863 (40.4%)	185,334 (2.9%)
Montana	5,068 (61.5%)	8,234 (35.4%)	23,289 (0.4%)
Nebraska	13,413 (71.1%)	18,855 (40.3%)	46,760 (0.7%)
Nevada	23,924 (92.9%)	25,757 (37.0%)	69,699 (1.1%)
New Hampshire	14,522 (59.1%)	24,585 (40.7%)	60,338 (0.9%)
New Jersey	36,795 (72.1%)	51,046 (40.6%)	125,796 (2.0%)
New Mexico	10,064 (64.7%)	15,565 (33.4%)	46,587 (0.7%)
New York	28,079 (77.9%)	36,029 (27.9%)	129,154 (2.0%)
North Carolina	68,341 (63.5%)	107,557 (27.9%)	129,154 (3.9%)
North Dakota	5,233 (62.3%)	8,394 (45.8%)	18,325 (0.3%)
Ohio	62,248 (78.5%)	79,250 (51.2%)	154,643 (2.4%)
Oklahoma	39,193 (65.8%)	59,570 (41.9%)	142,088 (2.2%)
Oregon	35,084 (82.5%)	42,539 (57.7%)	73,684 (1.1%)
Pennsylvania	65,328 (70.1%)	93,168 (40.5%)	230,153 (3.6%)
Rhode Island	3,368 (76.8%)	4,383 (61.3%)	7,150 (0.1%)
South Carolina	23,786 (59.0%)	16,550 (34.4%)	117,118 (1.8%)
South Dakota	5,787 (43.8%)	13,217 (49.2%)	26,888 (0.4%)
Tennessee	77,359 (80.6%)	95,938 (52.2%)	183,762 (2.9%)
Texas	359,852 (76.3%)	471,705 (48.4%)	974,475 (15.1%)
Utah	38,694 (82.8%)	46,749 (54.6%)	85,572 (1.3%)
Vermont	1,177 (71.7%)	1,642 (12.8%)	12,823 (0.2%)
Virginia	59,396 (67.2%)	88,378 (53.5%)	165,236 (2.6%)
Washington	70,512 (78.9%)	89,314 (59.9%)	149,184 (2.3%)
West Virginia	11,264 (35.9%)	31,351 (38.2%)	81,988 (1.3%)
Wisconsin	26,173 (64.6%)	40,518 (50.7%)	79,964 (1.2%)
Wyoming	1,787 (37.7%)	4,745 (40.4%)	11,759 (0.2%)
Total	2,004,752 (70.5%)	2,844,428 (44.2%)	6,440,043 (100.0%)

Source: Pre-AC file, Post-AC file, January 2009 ACS MAFX, the 2010 Census Combo File, 2010 Census Unedited File, and GeoStan Imputations File.
Note: Percentages may not sum to 100 due to rounding.

Table A-4. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: GIS Coordinates Actions by State

State	Add Actions (New Geocodes)		Change Actions (Corrected Geocodes)		Total GIS Coordinate based actions	
Alabama	1,907	(0.5%)	21,882	(0.7%)	23,789	(0.6%)
Arkansas	13,035	(3.7%)	120,205	(3.6%)	133,240	(3.6%)
California	13,299	(3.8%)	211,341	(6.3%)	224,640	(6.1%)
Colorado	3,558	(1.0%)	133,236	(4.0%)	136,794	(3.7%)
Florida	22,838	(6.5%)	67,118	(2.0%)	89,956	(2.4%)
Idaho	12,160	(3.5%)	41,637	(1.2%)	53,797	(1.5%)
Illinois	16,836	(4.8%)	190,928	(5.7%)	207,764	(5.6%)
Indiana	***30,895***	***(8.9%)***	***369,356***	***(11.0%)***	***400,251***	***(10.8%)***
Kansas	7,729	(2.2%)	108,702	(3.2%)	116,431	(3.1%)
Kentucky	***5,638***	***(1.6%)***	***303,306***	***(9.0%)***	***308,944***	***(8.3%)***
Louisiana	14,418	(4.1%)	84,951	(2.5%)	99,369	(2.7%)
Michigan	1,871	(0.5%)	18,314	(0.5%)	20,185	(0.5%)
Minnesota	340	(0.1%)	7,436	(0.2%)	7,776	(0.2%)
North Carolina	***95,929***	***(27.5%)***	***580,300***	***(17.3%)***	***676,229***	***(18.2%)***
North Dakota	2	(0.0%)	867	(0.0%)	869	(0.0%)
Ohio	7,358	(2.1%)	47,887	(1.4%)	55,245	(1.5%)
Oregon	10,555	(3.0%)	84,402	(2.5%)	94,957	(2.6%)
Pennsylvania	39,666	(11.4%)	198,546	(5.9%)	238,212	(6.4%)
South Carolina	4,716	(1.4%)	104,063	(3.1%)	108,779	(2.9%)
Texas	2,668	(0.8%)	5,261	(0.2%)	7,929	(0.2%)
Vermont	12,297	(3.5%)	71,384	(2.1%)	83,681	(2.3%)
Virginia	***26,099***	***(7.5%)***	***507,554***	***(15.1%)***	***533,653***	***(14.4%)***
Washington	0	(0.0%)	53,973	(1.6%)	53,973	(1.5%)
Wisconsin	5,095	(1.5%)	26,135	(0.8%)	31,230	(0.8%)
Total	348,909	(100.0%)	3,358,784	(100.0%)	3,707,693	(100.0%)

Source: January 2009 ACS MAFX and the 2010 Census Combo File.
Note: Percentages may not sum to 100 due to rounding.

Table A-5. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: GIS Coordinates Actions by regular AC Action

AC results	Add Actions (New Geocodes)	Change Actions (Corrected Geocodes)	Total GIS Coordinate based actions
Matched to AC add	1,599	6,294	7,893
	(0.5%)	(0.2%)	(0.2%)
Verify	73,886	2,702,256	2,776,142
	(21.2%)	(80.5%)	(74.9%)
Change	24,042	211,931	235,973
	(6.9%)	(6.3%)	(6.4%)
Move	4,033	99,366	103,399
	(1.2%)	(3.0%)	(2.8%)
Delete (verified)	158,748	103,669	262,417
	(45.5%)	(3.1%)	(7.1%)
Single Delete	3	11	14
	(0.0%)	(0.0%)	(0.0%)
Duplicate	30,395	24,545	54,940
	(8.7%)	(0.7%)	(1.5%)
Nonresidential	19,333	30,951	50,284
	(5.5%)	(0.9%)	(1.4%)
Uninhabitable	1,793	6,052	7,845
	(0.5%)	(0.2%)	(0.2%)
No regular AC action	35,077	173,709	208,786
	(10.1%)	(5.2%)	(5.6%)
Total records	348,909	3,358,784	3,707,693
	(100%)	(100%)	(100%)

Source: January 2009 ACS MAFX and the 2010 Census Combo File.
Note: Percentages may not sum to 100 due to rounding.

Table A-6. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: GIS Coordinates Actions by Large Block AC Action

AC results	Add Actions (New Geocodes)	Change Actions (Corrected Geocodes)	Total GIS Coordinate based actions
Matched to AC add	38	320	358
	(0.0%)	(0.0%)	(0.0%)
Verify	0	37	37
	(0.0%)	(0.0%)	(0.0%)
Change	2,987	20,250	23,237
	(0.9%)	(0.6%)	(0.6%)
Move	0	0	0
	(0.0%)	(0.0%)	(0.0%)
Delete (verified)	14,741	2,781	17,522
	(4.2%)	(0.1%)	(0.5%)
Single Delete	0	0	0
	(0.0%)	(0.0%)	(0.0%)
Duplicate	82	32	114
	(0.0%)	(0.0%)	(0.0%)
Nonresidential	39	39	78
	(0.0%)	(0.0%)	(0.0%)
Uninhabitable	3	62	65
	(0.0%)	(0.0%)	(0.0%)
No Large Block action	331,019	3,335,263	169,280
	(94.9%)	(99.3%)	(98.9%)
Total records	348,909	3,358,784	3,707,693
	(100.0%)	(100.0%)	(100.0%)

Source: January 2009 ACS MAFX and the 2010 Census Combo File.
Note: Percentages may not sum to 100 due to rounding.

Table A-7. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: ACS TOI Actions by State

State	ACS TOI Change actions		ACS TOI Delete actions		ACS TOI Nonresidential actions	
Alabama	8,037	(2.1%)	2,865	(3.1%)	486	(2.3%)
Alaska	6,367	(1.6%)	451	(0.5%)	223	(1.1%)
Arizona	7,226	(1.8%)	1,629	(1.7%)	376	(1.8%)
Arkansas	6,996	(1.8%)	1,825	(1.9%)	276	(1.3%)
California	10,230	(2.6%)	4,359	(4.6%)	1,259	(6.1%)
Colorado	4,301	(1.1%)	963	(1.0%)	289	(1.4%)
Connecticut	918	(0.2%)	659	(0.7%)	124	(0.6%)
Delaware	2,191	(0.6%)	343	(0.4%)	67	(0.3%)
District of Columbia	170	(0.0%)	315	(0.3%)	63	(0.3%)
Florida	7,397	(1.9%)	5,049	(5.4%)	774	(3.7%)
Georgia	8,849	(2.3%)	3,466	(3.7%)	576	(2.8%)
Hawaii	2,473	(0.6%)	484	(0.5%)	116	(0.6%)
Idaho	3,726	(1.0%)	420	(0.4%)	160	(0.8%)
Illinois	9,959	(2.5%)	4,387	(4.7%)	739	(3.6%)
Indiana	4,512	(1.2%)	1,878	(2.0%)	357	(1.7%)
Iowa	2,866	(0.7%)	1,544	(1.6%)	263	(1.3%)
Kansas	4,026	(1.0%)	1,360	(1.5%)	281	(1.4%)
Kentucky	7,161	(1.8%)	1,957	(2.1%)	318	(1.5%)
Louisiana	5,635	(1.4%)	2,733	(2.9%)	417	(2.0%)
Maine	17,401	(4.5%)	905	(1.0%)	279	(1.3%)
Maryland	1,638	(0.4%)	1,341	(1.4%)	237	(1.1%)
Massachusetts	2,926	(0.7%)	1,256	(1.3%)	213	(1.0%)
Michigan	11,811	(3.0%)	3,070	(3.3%)	658	(3.2%)
Minnesota	20,480	(5.2%)	2,483	(2.6%)	536	(2.6%)
Mississippi	7,194	(1.8%)	1,879	(2.0%)	371	(1.8%)
Missouri	11,882	(3.0%)	2,656	(2.8%)	558	(2.7%)
Montana	6,778	(1.7%)	430	(0.5%)	221	(1.1%)
Nebraska	5,356	(1.4%)	665	(0.7%)	249	(1.2%)
Nevada	1,261	(0.3%)	568	(0.6%)	133	(0.6%)
New Hampshire	6,410	(1.6%)	522	(0.6%)	113	(0.5%)
New Jersey	2,636	(0.7%)	2,068	(2.2%)	383	(1.8%)
New Mexico	6,183	(1.6%)	900	(1.0%)	309	(1.5%)
New York	26,264	(6.7%)	6,328	(6.7%)	1,569	(7.6%)
North Carolina	13,681	(3.5%)	3,126	(3.3%)	562	(2.7%)
North Dakota	2,815	(0.7%)	395	(0.4%)	121	(0.6%)
Ohio	4,827	(1.2%)	2,578	(2.7%)	492	(2.4%)
Oklahoma	13,870	(3.5%)	2,264	(2.4%)	517	(2.5%)
Oregon	2,542	(0.7%)	673	(0.7%)	176	(0.8%)
Pennsylvania	31,857	(8.2%)	4,051	(4.3%)	1,103	(5.3%)
Rhode Island	439	(0.1%)	298	(0.3%)	59	(0.3%)
South Carolina	5,101	(1.3%)	1,585	(1.7%)	338	(1.6%)
South Dakota	3,300	(0.8%)	408	(0.4%)	146	(0.7%)
Tennessee	5,484	(1.4%)	2,112	(2.3%)	437	(2.1%)
Texas	35,081	(9.0%)	7,614	(8.1%)	1,864	(9.0%)
Utah	2,851	(0.7%)	533	(0.6%)	140	(0.7%)
Vermont	8,887	(2.3%)	597	(0.6%)	136	(0.7%)
Virginia	6,999	(1.8%)	1,441	(1.5%)	326	(1.6%)
Washington	5,176	(1.3%)	1,174	(1.3%)	369	(1.8%)
West Virginia	8,480	(2.2%)	1,017	(1.1%)	355	(1.7%)
Wisconsin	6,557	(1.7%)	1,870	(2.0%)	499	(2.4%)
Wyoming	1,660	(0.4%)	266	(0.3%)	116	(0.6%)
Total	390,867	(100.0%)	93,760	(100.0%)	20,749	(100.0%)

Source: January 2009 ACS MAFX and the 2010 Census Combo File.
Note: Percentages may not sum to 100 due to rounding.

Table A-8. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: ACS TOI Actions by regular AC Action

AC results	ACS TOI Change actions	ACS TOI Delete actions	ACS TOI Nonresidential actions
No regular AC Action	38,628	9,119	6,605
	(9.5%)	(8.8%)	(31.5%)
Matched to AC Add	1,803	94	558
	(0.5%)	(0.1%)	(2.8%)
Verify	124,131	14,869	3,006
	(31.9%)	(16.0%)	(14.7%)
Change	113,710	9,767	2,549
	(29%)	(11%)	(12%)
Move	4,828	823	98
	(1.2%)	(0.9%)	(0.5%)
Delete (verified)	63,140	51,830	3,927
	(16.3%)	(56.0%)	(19.0%)
Single Delete	2	0	0
	(0.0%)	(0.0%)	(0.0%)
Duplicate	39,756	4,919	1,068
	(10.1%)	(5.1%)	(4.9%)
Nonresidential	1,968	1,353	2,664
	(0.5%)	(1.4%)	(12.9%)
Uninhabitable	2,901	986	274
	(0.7%)	(1.1%)	(1.3%)
Total records	390,867	93,760	20,749
	(100.0%)	(100.0%)	(100.0%)

Source: January 2009 ACS MAFX and the 2010 Census Combo File.
Note: Percentages may not sum to 100 due to rounding.

Table A-9. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: ACS TOI Actions by Large Block AC Action

AC results	ACS TOI Change actions	ACS TOI Delete actions	ACS TOI Nonresidential actions
No Large Block AC Action	389,425	92,950	20,704
	(99.6%)	(99.1%)	(99.8%)
Matched to AC Add	42	11	11
	(0.0%)	(0.0%)	(0.1%)
Verify	62	2	0
	(0.0%)	(0.0%)	(0.0%)
Change	857	184	18
	(0.2%)	(0.2%)	(0.1%)
Move	3	0	0
	(0.0%)	(0.0%)	(0.0%)
Delete (verified)	435	613	14
	(0.1%)	(0.7%)	(0.1%)
Single Delete	0	0	0
	(0.0%)	(0.0%)	(0.0%)
Duplicate	31	0	0
	(0.0%)	(0.0%)	(0.0%)
Nonresidential	9	0	2
	(0.0%)	(0.0%)	(0.0%)
Uninhabitable	3	0	0
	(0.0%)	(0.0%)	(0.0%)
Total records	390,867	93,760	20,749
	(100.0%)	(100.0%)	(100.0%)

Source: January 2009 ACS MAFX and the 2010 Census Combo File.
Note: Percentages may not sum to 100 due to rounding.

Table A-10. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: STARS/MTdb matching records by source of AC Action

AC results	Regular AC	Large Block AC
No AC Action (of column type)	9,421,019	117,558,406
	(7.9%)	(98.4%)
New AC Add	262,848	2,317
	(0.2%)	(0.0%)
Matched to AC Add	2,694,246	100,229
	(2.3%)	(0.1%)
Verify	85,892,227	17,749
	(71.9%)	(0.0%)
Change	11,097,437	1,554,786
	(9.3%)	(1.3%)
Move	4,814,804	1,096
	(4.0%)	(0.0%)
Delete (verified)	3,721,248	179,544
	(3.1%)	(0.2%)
Single Delete	371	0
	(0.0%)	(0.0%)
Duplicate	1,014,093	7,973
	(0.8%)	(0.0%)
Nonresidential	342,357	1,098
	(0.3%)	(0.0%)
Uninhabitable	166,621	4,073
	(0.1%)	(0.0%)
Total records	119,427,271	119,427,271
	(100.0%)	(100.0%)

Source: January 2009 ACS MAFX and the 2010 Census Combo File.
Note: Percentages may not sum to 100 due to rounding.