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

February 6, 2013

2010 CENSUS PLANNING MEMORANDA SERIES

No. 252

MEMORANDUM FOR The Distribution List

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

Subject: 2010 Census Evaluation of Address Frame Accuracy and Quality

Attached is the 2010 Census Evaluation of Address Frame Accuracy and Quality. 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 Nancy R. Johnson at (301) 763-3639, Kathleen Kephart at (301) 763-8891 or Kevin Shaw at (301) 763-1851.

Attachment

2010 Census Evaluation of Address Frame Accuracy and Quality

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

Nancy Johnson and Kathleen Kephart

Decennial Statistical Studies Division

This page intentionally left blank.

Table of Contents

List of Tables

This page intentionally left blank

Executive Summary

This report provides results from the 2010 Census Evaluation of Address Frame Accuracy and Quality. The Decennial Statistical Studies Division conducted this evaluation as part of the 2010 Census Program for Evaluations and Experiments to evaluate the accuracy of the address frame after the 2010 Census Address Canvassing operation and after completion of all 2010 Census operations. Specifically, this evaluation used results from the 2010 Census Coverage Measurement operation and a special supplemental field operation to estimate the number and percent of housing units correctly added (and added-in-error) and correctly deleted (and deleted-in-error) by census operations. In addition, this evaluation examined the universe of addresses coded as "missing" from the census by the Census Coverage Measurement operation to determine if the census included them as valid housing units, but incorrectly geocoded them to a 2010 Census collection block outside of the Census Coverage Measurement geographic search area. This geocoding error analysis included a second supplemental field operation.

This evaluation was based on two Census 2000 studies:

1. "Analysis of Deleted and Added Housing Units in Census 2000 Measured by the Accuracy and Coverage Evaluation" (Smith et al., 2003), and
2. "An Assessment of Addresses on the Master Address File 'Missing' in the Census or Geocoded to the Wrong Collection Block" (Ruhnke, 2003).

This evaluation expands on the work conducted by Smith et al. (2003) by including a component to estimate the number and percent of addresses that the 2010 Census Address Canvassing operation correctly added (and added-in-error) and correctly deleted (and deleted-in-error). The Smith et al. evaluation (2003) did not produce similar estimates for either Census 2000 Block Canvassing or Census 2000 Address Listing.

This evaluation addresses the following research questions:

How accurate was the address frame pre-Address Canvassing, post-Address Canvassing, and after final census operations? How can we improve address frame quality?

Since the accuracy and quality of the address frame is a broad topic that can include many components, this evaluation focused on changes and improvements to the census address inventory and narrowed the analysis to answer six operational study questions, as described in the following three paragraphs.

The "address frame" during the course of census operations is dynamic. The Pre-Address Canvassing address inventory, serving as the basis for the 2010 Census, is called the Master Address File. Measuring the accuracy and quality of this address frame was outside the scope of this evaluation due to timing, resources, and limitations in the difficulty of evaluating this universe. Our analysis provides documentation of the changes that were made to the address frame in the course of conducting the 2010 Census, which cannot be interpreted as a direct measure of accuracy of the address frame, but can give an indication of what updates were required to get the address frame "ready" for the 2010 Census.

While some of the records on the Master Address File represent valid living quarters, other records represent duplicate units, nonresidential addresses, nonexistent units, uninhabitable units, or other types of invalid records. A subset of the Master Address File records that were the most likely to be potential valid living quarters went into the Address Canvassing operation for Listers to validate or correct. In addition, the Listers added addresses missing from this list. In the post-Address Canvassing list, this evaluation measured the accuracy of actions taken during Address Canvassing and, specifically, answered operational study questions one and two below.

After Address Canvassing, a Group Quarters Validation operation made further changes and enhancements to the census address inventory. This inventory then served as the basis for the Universe Control and Management system. The Universe Control and Management system tracked the flow of data to and from the census enumeration operations. At the completion of all census operations, the final status of living quarters is reflected on the Census Unedited File. Using the Census Unedited File, this evaluation measured the accuracy of housing units by answering operational study questions three to six below.

The six operational study questions are as follows:

1. **By source/operation, what proportion of Address Canvassing "Adds" were correctly added and erroneously added, according to their Census Coverage Measurement status?**
2. **By source/operation, what proportion of Address Canvassing deleted/duplicated units were correctly deleted and erroneously deleted, according to their Census Coverage Measurement status?**
3. **By source/operation, what proportion of the post-Address Canvassing "Adds" were correctly added and erroneously added, according to their final Census Coverage Measurement status?**
4. **By source/operation, what proportion of the post-Address Canvassing deleted/duplicated units were correctly deleted and erroneously deleted, according to their final Census Coverage Measurement status?**
5. **What was the total estimated percentage of census addresses geocoded to the incorrect 2010 Census collection block?**
6. **Did the geocoding error estimate vary by type of enumeration area or by census region?**

As a by-product of the analysis, this evaluation also investigated whether the added and deleted housing units were geographically clustered in the 2010 Census Coverage Measurement sample block clusters or evenly distributed throughout. This additional analysis is included within the results for each of the operational study questions one through four.

The table below summarizes the results of the operational study questions one through four, providing estimates for the 2010 Census Address Canvassing deleted/duplicated addresses, the Address Canvassing added addresses, the Census Unedited File deleted/duplicated addresses, and the Census Unedited File added addresses. The table also compares the 2010 Census estimates to the Census 2000 estimates of addresses deleted/duplicated on the Decennial Master Address File and of addresses added to the Hundred Percent Census Unedited File. The post 2010 Census Address Canvassing file and the Census Unedited File are analogous to the Census 2000 Decennial Master Address File and the Hundred Percent Census Unedited File, respectively. As in Census 2000, the Census Unedited File has the final census status of the records.

2010 CPEX* Address Frame Accuracy and Quality:
Census 2000 and 2010 Census Comparison –
Estimates of Housing Units Deleted/Duplicated and Added-in-Error

Types of Actions	Census 2000	2010 Census	
	DMAF/HCUF* Weighted Count (Percent) (SE[1,2])	Address Canvassing Weighted Count (Percent) (SE[1])	Census Unedited File Weighted Count (Percent) (SE[1])
Total Delete/Duplicate Housing Units	**8,536,752**	**18,445,131**	**4,850,528**
Correctly Deleted.....	7,309,409	17,658,837	3,599,162
	(85.6)	(95.7)	(74.2)
	(n/a)	(0.6)	(1.8)
Deleted-in-Error …….	1,227,343	786,294	1,251,366
	(14.4)	(4.3)	(25.8)
	(n/a)	(0.6)	(1.8)
Total Added Housing Units	**3,857,381**	**10,585,463**	**3,601,110**
Correctly Added ……..….	3,235,099	8,853,529	2,867,070
	(83.9)	(83.6)	(79.6)
	(n/a)	(1.3)	(2.5)
Added-in-Error …….....	622,282	1,731,934	734,040
	(16.1)	(16.4)	(20.4)
	(n/a)	(1.3)	(2.5)

[1] Standard Error.
[2] Standard Errors for Census 2000 estimates are not available.
*Census Program for Experiment and Evaluations, Decennial Master Address File , Hundred Percent Census Unedited File
Source: Smith et al., 2003, 2010 Census Address Canvassing Deletes/Duplicates Analysis File, 2010 Census Address Canvassing Adds Analysis File, 2010 Census Unedited File Delete/Duplicate Analysis File and 2010 Census Unedited File Adds Analysis File.

Summaries of the results for each of the six operational study questions follow:

1. By source/operation, what proportion of Address Canvassing "Adds" were correctly added and erroneously added, according to their Census Coverage Measurement status?

The 2010 Census Address Canvassing operation correctly added about 83.6 percent of the addresses having an add action and added about 16.4 percent of the add records in error. Of the

addresses added-in-error, about 11.9 percent were not a housing unit or did not exist on April 1, 2010 (Census Day) according to the Census Coverage Measurement results.

This evaluation results suggest that the Address Canvassing addresses added-in-error were clustered geographically and were not distributed evenly throughout the Census Coverage Measurement sample block clusters. About 46 percent of the sample block clusters had at least one Address Canvassing add action. Of the 2,803 sample block clusters with an added address, about 50 percent had no addresses added-in-error.

2. By source/operation, what proportion of Address Canvassing deleted/duplicated units were correctly deleted and erroneously deleted, according to their Census Coverage Measurement status?

Address Canvassing correctly deleted most (95.7 percent) of the addresses having a delete or duplicate action and deleted-in-error 4.3 percent of the deleted/duplicated addresses. This represented a weighted count of 786,294 addresses deleted-in-error and about 0.5 percent of all records processed in the Address Canvassing operation.

This evaluation results suggest that the Address Canvassing addresses deleted-in-error were clustered geographically and were not distributed evenly throughout the Census Coverage Measurement sample block clusters. About 68 percent of the sample block clusters had at least one Address Canvassing delete/duplicate action. Of these sample block clusters with deleted/duplicated address, about 80 percent did not contain any addresses deleted-in-error.

3. By source/operation, what proportion of the post-Address Canvassing "Adds" were correctly added and erroneously added, according to their final Census Coverage Measurement status?

The Census Coverage Measurement results confirm that the census operations correctly added 79.6 percent of the Census Unedited File Adds. This means census operations added-in-error 20.4 percent of the adds. According to the Census Coverage Measurement, about 6.0 percent of the Adds did not exist as a housing unit on Census Day. Although the rate of address records added-in-error was higher than the rate (16.1 percent) found in the Smith et al. (2003), the Decennial Statistical Studies Division could not determine whether this difference was statistically significant because standard errors were not available from the Census 2000 evaluation. However, the ratio of total added addresses in the 2010 Census to valid housing units of 2.5 percent was lower than the Census 2000 ratio of 3.3 percent.

This evaluation results suggest that the Census Unedited File addresses added-in-error were clustered geographically and were not distributed evenly throughout the Census Coverage Measurement sample block clusters. About 18.8 percent of the 6,148 sample block clusters had at least one added address. Of these sample block clusters with an added address, about 55.9 percent had no addresses added-in-error.

4. By source/operation, what proportion of the post-Address Canvassing deleted/duplicated units were correctly deleted and erroneously deleted, according to their final Census Coverage Measurement status?

The census operations correctly deleted most (74.2 percent) of the Census Unedited File Deletes/Duplicates and deleted-in-error 25.8 percent of the deleted/duplicated records. The records deleted-in-error represented a weighted count of 1,251,366 addresses that may have been omitted by the census. This estimate was nearly identical to the weighted estimate of 1.2 million housing units found by Smith et al. (2003) to be deleted-in-error in Census 2000.

This evaluation results suggest that the Census Unedited File addresses deleted-in-error were clustered geographically and were not distributed evenly throughout the Census Coverage Measurement sample block clusters. About 48.5 percent of the 6,148 sample block clusters had at least one Census Unedited File delete/duplicate address. Of these sample block clusters with a deleted/duplicated address, about 61.6 percent had no addresses deleted-in-error.

5. What was the total estimated percentage of census addresses geocoded to the incorrect 2010 Census collection block?

The Census Coverage Measurement limited the matching of addresses to within the sample block cluster and one ring of blocks surrounding the cluster. By matching addresses in a larger geographic area than the one ring of surrounding blocks, this evaluation found additional addresses that were missing-from-census, but were in the 2010 Census in the incorrect block. The analysis assigned blocks in the search area to rings of one kilometer, three kilometers, and five kilometers surrounding the block cluster. As a result, the Decennial Statistical Studies Division was able to determine whether geocoding error occurred more frequently in blocks closest to the subject block (within one kilometer) or in blocks farther away from the subject block (five kilometers).

The estimated percent of census addresses geocoded to the incorrect 2010 Census collection block is 1.5 percent (standard error is 0.6 percent) in the expanded search area up to five kilometers from the block cluster. This evaluation found a geocoding error rate of 1.4 percent within the cluster and one surrounding ring of blocks. An additional 0.1 percent of units were found misgeocoded within the one kilometer buffer of the block cluster. Beyond one kilometer, not enough cases were found to produce a statistically valid estimate, with only 28 unweighted cases. The 2010 Census estimate of geocoding error is lower than the estimate of 4.8 percent (standard error is 0.3 percent) found by Ruhnke (2003).

6. Did the geocoding error estimate vary by type of enumeration area or by census region?

The geocoding error rate is the same as the national average for the Mailout/Mailback enumeration areas (1.5 percent) and higher for Update/Leave areas, at 1.9 percent.

The geocoding error is highest in the South at 2.3 percent while the Midwest has the lowest rate at 0.7 percent. The Northeast and the West had similar error rates at 1.2 percent and 1.1 percent, respectively.

Based on results from this evaluation, the authors recommend the following:

- **Conduct research into the characteristics and predictors of collection blocks with added and deleted addresses.**

  Understanding the characteristics/predictors of change to the census address inventory is an important step in maintaining an accurate and high quality address frame. Some characteristics are well known, such as areas that experience new construction, but predictors of changes in the address inventory have received less study. Research using the 2010 Census Planning Database, Administrative Records, and information from the Partnership Program, as well as data from 2020 Census field tests can provide information to predict future changes. Areas with added and deleted addresses are likely to be spatially/geographically clustered. This finding is consistent with Boies (2012), and has important implications for future decisions on targeting certain blocks for listing operations.

- **Conduct research into the characteristics and predictors of addresses added-in-error and deleted-in-error.**

  Understanding where and why the census operations make errors in either adding or deleting addresses is important for an accurate census. The Census Bureau should conduct additional research into the addresses that the census added and deleted-in-error to determine characteristics of these addresses or of their blocks. Research using the 2010 Census Planning Database, Administrative Records, the Partnership Program and the 2020 Census field tests could provide information on specific addresses or blocks that need targeted training, procedures, or updating.

- **Weigh the costs and benefits of further decreasing geocoding error.**

  The results of this evaluation indicate that the geocoding error was significantly less than in Census 2000 and the authors recommend that the Census Bureau examine the costs and benefits associated with further decreasing geocoding error. Given a geocoding error rate of only 1.5 percent, it may not be worth implementing large-scale field operations only to improve geocoding. However, several low cost options may be available to continue to improve geocoding on the Master Address File. According to Tomaszewski (*Forthcoming 2013*), the majority of the Geographic Information Systems geocodes collected from local governments were accurate. Continued collection of these Geographic Information Systems data may lead to a smaller number of missing, ungeocoded, and misgeocoded units in the address frame, which translates into a higher quality address frame (reduced undercoverage and increased accuracy).

In addition to analyzing the costs and benefits associated with improving geocoding error, the authors also believe that the current Census Coverage Measurement search area is sufficient to measure geocoding error. Since Census Coverage Measurement already matches records within one ring of surrounding blocks it may not be necessary to perform a separate evaluation, independent of the 2020 Census post enumeration survey measure. This assumes producing this estimate is in line with the goals and design of the 2020 Census post enumeration survey program.

- **Fully Integrate the housing unit added- and deleted-in-error statistics into the 2020 Census Coverage Measurement housing unit studies.**

  In this evaluation, in order to calculate the housing unit added- and deleted-in-error statistics for the final 2010 Census, the 2010 Census Coverage Measurement results were heavily leveraged. Additionally, very similar headquarters and field procedures were used to determine the 2010 Census Address Canvassing added- and deleted-in-error statistics presented here. With these facts, combined with the knowledge that 2010 Census Coverage Measurement staff have and continue to conduct very similar work, it may be more economical for the Census Bureau to integrate these added- and deleted-in-error statistics, both post-listing and final census, into the 2020 Census post enumeration survey program. With increased funding and priority over the 2020 Census Evaluation program, the 2020 Census post enumeration survey program may be better suited to conduct the fieldwork and subsequent analyses in a timely manner; without the need to design, plan, and execute costly supplemental field operations.

1 Introduction

1.1 Scope

The purpose of this evaluation was to estimate the accuracy of the address frame after both the 2010 Census Address Canvassing (AC) operation and after completion of all 2010 Census operations. Using results from the 2010 Census Coverage Measurement (CCM) program and a special supplemental field operation, this evaluation analyzed housing units erroneously added and erroneously deleted by census operations. In addition, the Decennial Statistical Studies Division (DSSD) examined the addresses coded as "missing" from the census by CCM to determine if the census included them as valid housing units, but incorrectly geocoded them to a collection block outside of the CCM geographic search area.

This evaluation addressed the following research questions:

How accurate was the address frame pre-AC, post-AC, and after final census operations? How can we improve address frame quality?

Since the accuracy and quality of the address frame is a broad topic that can include many components, this evaluation focused on changes and improvements to the census address inventory and narrowed the analysis to answer six operational study questions, as described in the following three paragraphs.

The address frame during the course of census operations is dynamic. The pre-AC address inventory, serving as the basis for the 2010 Census, is called the Master Address File (MAF). Measuring the accuracy and quality of this address frame was outside the scope of this evaluation due to timing, resources, and limitations in the difficulty of evaluating this universe. Our analysis provides documentation of the changes that were made to the address frame in the course of conducting the 2010 Census, which cannot be interpreted as a direct measure of accuracy of the address frame, but can give an indication of what updates were required to get the address frame "ready" for the 2010 Census.

While some of the records on the MAF represent valid living quarters, other records represent duplicate units, nonresidential addresses, nonexistent units, uninhabitable units, or other types of invalid records. A subset of the MAF records that were the most likely to be potential valid living quarters went into the AC operation for Listers to validate or correct. In addition, the Listers added addresses missing from this list. In the post-AC list, this evaluation measured the accuracy of actions taken during AC and, specifically, answered operational study questions one and two below.

After AC, a Group Quarters Validation (GQV) operation made further changes and enhancements to the census address inventory. This inventory then served as the basis for the Universe Control and Management (UCM) system. The UCM monitored the flow of data to and from the census enumeration operations. At the completion of all census operations, the final status of living quarters is reflected on the Census Unedited File (CUF). In the CUF, this

evaluation measured the accuracy of housing units by answering the third through sixth operational study questions below.

The six operational study questions are as follows:

1. By source/operation, what proportion of Address Canvassing "Adds" were correctly added and erroneously added, according to their Census Coverage Measurement status?
2. By source/operation, what proportion of Address Canvassing deleted/duplicated units were correctly deleted and erroneously deleted, according to their Census Coverage Measurement status?
3. By source/operation, what proportion of the post-Address Canvassing "Adds" were correctly added and erroneously added, according to their final Census Coverage Measurement status?
4. By source/operation, what proportion of the post-Address Canvassing deleted/duplicated units were correctly deleted and erroneously deleted, according to their final Census Coverage Measurement status?
5. What was the total estimated percentage of census addresses geocoded to the incorrect 2010 Census collection block?
6. Did the geocoding error estimate vary by type of enumeration area or by census region?

1.2 Intended Audience

This report is intended for managers and staff involved in planning and implementing the 2020 Census and for other stakeholders interested in the accuracy and quality of the Census Bureau address frame.

2 Background

This evaluation included an analysis of deleted and added addresses in the 2010 Census AC operation. This evaluation repeated and expanded on two Census 2000 studies: 1) an analysis of deleted and added housing units, and 2) an assessment of addresses "missing" or geocoded to the wrong block. The Census 2000 study on deleted and added housing units did not include a component for either the Address Listing (AL) operation or the Block Canvassing (BC) operation.

2.1 Background on Previous Studies

Highlights from the two Census 2000 studies are below:

1. Census 2000 Analysis of Deleted and Added Housing Units

The Census 2000 Evaluation O.19 titled, "Analysis of Deleted and Added Housing Units in Census 2000 Measured by the Accuracy and Coverage Evaluation" (Smith et al., 2003), examined the changes in the census housing unit inventory between the creation of the January 2000 Decennial Master Address File (DMAF) and the Hundred Percent Census Unedited File (HCUF). This evaluation defined census deleted addresses as units that were on the January

2000 DMAF but not on the HCUF. Similarly, the census adds were defined as units that were on the HCUF but not on the January 2000 DMAF. Results are in the table below, all numbers in the report are weighted estimates based on a sample. All table captions contain a heading for the Census Program for Evaluations and Experiments (CPEX).

Table 1. CPEX Address Frame Accuracy and Quality: Census 2000 Estimates of Housing Units Deleted and Added-in-Error

Types of Actions	Weighted Count	Weighted Percent
Total Deleted Units	8,536,752	100.0
Correctly Deleted	7,309,409	85.6
Deleted-in-Error	1,227,343	14.4
Total Added Units	3,857,381	100.0
Correctly Added	3,235,099	83.9
Added-in-Error	622,282	16.1

Source: Smith et al., 2003.

In Census 2000 on the HCUF, there were 115,904,641 total housing units in the nation (Woodward and Damon, 2001). The total number of deleted units was 7.4 percent, while the total number of added units was 3.3 percent of the 115.9 million housing units in the nation. Census 2000 operations correctly deleted most of the deleted units and correctly added most of the added units.

Of the correctly deleted housing units, 63.1 percent were not housing units or did not exist on April 1, 2000. The Kill Process[1] (61.3 percent) and the Housing Unit Unduplication Operation[2] (29.3 percent) accounted for 90.7 percent of the 1.2 million erroneously deleted units. The Kill Process correctly deleted 89.7 percent of the units and the Housing Unit Unduplication Operation correctly deleted 64.0 percent of the units (Smith et al., 2003).

2. Census 2000 Assessment of Addresses 'Missing' or Geocoded to the Wrong Block

In Census 2000, the post enumeration survey that measured coverage error was called the Accuracy and Coverage Evaluation (A.C.E.).

[1] The Kill Process identified records that were not likely to uniquely represent housing units that existed on April 1, 2000, Census Day. One example of a category in this Kill Process that was excluded from the census was units that met the following conditions: no census form was returned for the unit, the unit was deleted in the Nonresponse Followup operation, and the unit was confirmed as a delete in the Coverage Improvement Followup operation.

[2] The Census 2000 Housing Unit Unduplication Operation consisted of two phases. The first phase identified potential duplicates on the DMAF through address and person matching algorithms. In the second phase, rules were developed and applied to exclude housing units that were likely to be duplicates of other census addresses. As a result of these rules, approximately 1.4 million housing units were excluded from the census.

The Census 2000 F.15 evaluation titled, "An Assessment of Addresses on the Master Address File 'Missing' in the Census or Geocoded to the Wrong Collection Block" (Ruhnke, 2003), used data from the Census 2000 A.C.E. to determine a geocoding error rate and whether the "missing" addresses were actually included in the census as housing units, but incorrectly geocoded to collection blocks outside the A.C.E. geographic search area. In addition, it investigated how many of the "missing" addresses were on the DMAF and the reasons they did not end up on the HCUF. Results were as follows. All numbers are weighted estimates based on a sample:

- About 4.8 percent of all housing units were geocoded to the incorrect collection block.

- About 1.3 million units coded as erroneously excluded from Census 2000 by A.C.E. were on the MAF but excluded due to the Census Bureau's rules for creating the DMAF and the HCUF.

2.2 Background on the Address Canvassing Operation

Prior to Census 2000 the BC operation consisted of a dependent listing of 51 percent of all non-water collection blocks in the U.S. (Burcham, 2002). The breakdown of actions in Census 2000 BC was:

- *Verify.....81,115,466 addresses*
- *Add.....6,389,271 addresses*
- *Delete.....5,146,320 addresses*
- *Address Corrected.....2,295,168 addresses*
- *Geographic Correction.....2,948,414 addresses*
- *Add and Verify.....107,402 addresses*

The AL operation was an independent listing of blocks in areas that the United States Postal Service Delivery Sequence Files (DSF) had insufficient coverage (post office boxes, rural route addresses). The AL workload was about 22 million records (Ruhnke, 2002).

The A.C.E. post enumeration survey was not used as a truth measure to estimate the rate of units added/deleted correctly from AL or BC. Instead, results from the A.C.E. were used to examine housing units added and deleted between the January 2000 DMAF and the HCUF.

Due to many methodological differences, the AL/BC operations are not truly comparable to AC. They are only mentioned in this report because they are the operations from Census 2000 most closely analogous to the AC operation. Some of the major differences include: AC was automated and BC and AL were paper based; AC was dependent listing and AL was independent; Quality Control (QC) was conducted by production staff for BC and AL while AC had an independent QC staff. For further discussion of the differences see Holland (2012).

Address Canvassing replaced the AL and BC operations that occurred in Census 2000. The purpose of the AC operation was to:

- Update the Census Bureau's address list to ensure an accurate frame for the enumeration of the population.

- Validate submissions from the Local Update of Census Addresses (LUCA) program and to allow the Census Bureau to provide feedback to the local governments participating in the program.

During the operation, production Listers canvassed assigned blocks and looked for every potential living quarter. Listers compared the addresses, map features, and types of living quarters they observed on the ground to the Census Bureau's address list and updated the list, where appropriate, using hand-held computers with software designed specifically for the operation. In addition, the Listers electronically collected structure coordinates that located the living quarters on census maps. Quality Control Listers verified a sample of addresses from each Assignment Area and all deletes and duplicates identified by the production Lister. If a Quality Control Lister deleted a record not deleted by the production Lister or marked a record as a duplicate, a second Quality Control Lister validated the delete or the duplication during the Final Delete Verification phase of the operation.

The Large Block Address Canvassing operation was a successful mitigation strategy for canvassing addresses in blocks containing too many addresses for the hand-held computer to process in an acceptable time. This mitigation strategy was designed after problems encountered during the 2008 Census Dress Rehearsal. Instead of using a hand-held computer, the operation worked with existing software and processes originally designed for Census Bureau survey operations that are similar to AC.

Overall, the AC operation added 10,300,593 living quarters, excluding Puerto Rico. Of these added units, 6,149,446 were new adds and 4,151,147 matched to an existing record on the MAF. A total of 15,529,724 records were deleted with 3,968,495 units marked as duplicates (Address List Operations Implementation Team, 2011).

2.3 Background on the 2010 Census Coverage Measurement Operation

The purpose of the 2010 CCM program was to evaluate coverage in the 2010 Census in order to aide in future censuses, meaning 2020 Census and beyond. The CCM program was designed to measure the coverage of housing units and persons, excluding Group Quarters (GQs) and persons residing in GQs. The CCM program provided estimates of net coverage, showing undercount and over count, and components of census coverage including omissions and erroneous inclusions. Since the CCM was an evaluation, the results did not affect the 2010 Census (Viehdorfer, 2011). The CCM program includes several operations related to person coverage including Person Matching, the Person Interview, and Person Follow-up these operations and will not be discussed in this section since they are outside this evaluation's scope.

The CCM program is a large, complex survey conducted independently of the census. The CCM program consists of six sampling activities, five data collection activities, and three matching activities. An estimation operation follows these activities. The remaining discussion

summarizes the CCM activities of interest to this evaluation. For more information on the CCM survey design, refer to "The Design of the Coverage Measurement Program for the 2010 Census—Revision #1," memorandum #2010-B-07-R1 in the DSSD 2010 Census Coverage Measurement Memorandum Series.

The CCM sample was an area-based sample that consisted of groups of census blocks, called block clusters[3]. The CCM Listers independently listed all block clusters in the sample. Drawing an area sample created an independent measure of address coverage. Then, the CCM used its independent list as a frame for the Population sample (P sample). The CCM program also selected an Enumeration sample (E sample) from the CUF. The E sample consists of census housing units and person enumerations in housing units in the same sample areas as the P sample.

After the Independent Listing (IL), CCM computer-matched its address list to units on the initial UCM[4] file (as of January 2010) included in the CCM sample block clusters plus one ring of surrounding blocks. The results were loaded into a clerical matching software database called the Housing Unit Matching, Review, and Coding System (HUMaRCS). The National Processing Center (NPC) conducted the clerical matching and sent discrepancies to a field followup operation.

Since this Initial Housing Unit (IHU) matching operation occurred before the census housing unit inventory was final, a Final Housing Unit (FHU) matching operation occurred. The FHU Computer Processing phase of this operation used information from all previous CCM operations along with final census data from the CUF to assign flags to units that needed review in the clerical matching phase as follows:

- P-sample links to census UCM Deletes[5]
- Census UCM Adds in the sample block cluster[6]
- Census UCM Adds in the blocks surrounding the CCM sample block cluster
- E-sample units that were duplicates, where the primary was deleted from the census UCM file and is not on the CUF
- E-sample units linked to a unit that was no longer in the P sample.

In addition, the computer processing identified CCM and census units, including census GQs[7], with links to census units not on the CUF and un-linked those units. When the computer processing was complete, the Final HU Before Followup (BFU) Clerical Matching phase began.

In the Final HU BFU Clerical Matching process, the matching staff attempted to match all flagged units by address, housing unit description, or map spot location. In addition to

[3] Block clusters consisted of one or more contiguous blocks.
[4] The initial UCM identified the addresses eligible for the initial questionnaire printing.
[5] A census UCM delete was a housing unit on the Initial UCM used for IHU that was *not* on the CUF.
[6] A census UCM add was a housing unit *not* on the Initial UCM but was on the CUF.
[7] While CCM identified GQs in the listing, matching, and field operations, it did not specifically classify their enumeration status (i.e., correct enumeration or erroneous enumeration) and did not provide estimates for GQs.

identifying these matches between CCM and census units, the matching staff attempted to identify duplicate units:

- Between CCM P sample units within the CCM sample block clusters
- Between E sample units and other census units in the CCM sample block cluster and in its surrounding blocks

Unresolved Clerical Matching units went to the FHU Followup operation. In the FHU Followup operation, the CCM field staff collected information in an attempt to resolve the status of the addresses. They collected the following information:

- Census collection block number
- Information to determine if a housing unit existed at the address on April 1, 2010, Census Day (If not, why?)
- Information to determine whether or not units identified as possible matches between CCM and the census really are the same
- Information to determine whether or not units identified as possible duplicates really are the same

Results of the FHU Followup went to the FHU After Followup (AFU) clerical matching process. In the FHU AFU clerical matching process, staff used information from all previous operations to determine a match code that indicated both a final housing unit status and final match status. Any housing units that still remained unresolved after this step were handled using statistical techniques for missing data. See Bray and Viehdorfer (2012) for more information.

3 Methodology

3.1 Research Questions

The research questions for this evaluation were: How accurate was the address frame pre-AC, post-AC, and after the final census operation? How can we improve address frame quality?

Specific operational study questions are:

1. **By source/operation, what proportion of Address Canvassing "Adds" were correctly added and erroneously added, according to their Census Coverage Measurement status?**
2. **By source/operation, what proportion of Address Canvassing deleted/duplicated units were correctly deleted and erroneously deleted, according to their Census Coverage Measurement status?**

3. **By source/operation, what proportion of the post-Address Canvassing "Adds" were correctly added and erroneously added, according to their final Census Coverage Measurement status?**
4. **By source/operation, what proportion of the post-Address Canvassing deleted/duplicated units were correctly deleted and erroneously deleted, according to their final Census Coverage Measurement status?**
5. **What was the total estimated percentage of census addresses geocoded to the incorrect 2010 Census collection block?**
6. **Did the geocoding error estimate vary by type of enumeration area or by census region?**

3.2 Office of Management and Budget Clearance

The field work for this evaluation was approved by the Office of Management and Budget (OMB) on September 17, 2010, title: 2010 Census Coverage Measurement Final Housing Unit Followup and Address Frame Accuracy and Quality Evaluation (OMB Control Number 0607-0962, ICR Reference Number 201007-0607-002).

3.3 Schedule

This evaluation had two field operations and each field operation included a clerical matching activity before the field followup and after completion of the field followup. In addition, the NPC conducted research on the AC and the census/CUF deleted units. The schedule for these activities is in Table 2.

Table 2. 2010 CPEX Address Frame Accuracy and Quality: Schedule for Evaluation Activities

Activity	Actual Dates
Address Canvassing Delete/Duplicate Operation	
Before Followup Clerical Matching	05/24/10-06/18/10
Field Followup	03/23/11-04/13/11
After Followup Clerical Matching	06/27/11-06/29/11
Geocoding Error Operation	
Before Followup Clerical Matching	08/10/11-09/01/11
Field Followup	09/06/11-09/30/11
After Followup Clerical Matching	10/17/12-10/20/12
Address Canvassing Delete Research	09/01/11-09/21/11
Census Delete Research	12/01/11-01/17/12

Source: Internal schedule.

3.4 Data for Address Canvassing Adds and Delete/Duplicate Actions

The DSSD used the following files to create the AC Delete/Duplicate and Add Action Analysis files:

- CCM Sample Design File, Version 2 (Konicki, 2011) SDF.102
- CCM Sample Design File, Version 3 (Konicki, 2011) SDF.103
- DSSD Data File (Konicki, 2010a) SBCDATA.101
- Group Quarters Validation (GQV) State files (Lynch, 2009)
- DSSD Combo State Files (Ward, 2011)
- CCM Initial Housing Unit Independent List file hum_iladdfull.sas7bdat (Imel and Probst, 2012)
- CCM Initial Housing Unit Census Address List file hum_cenaddfull.sas7bdat (Imel and Probst, 2012)
- CCM Final Housing Unit Independent List file fhum_ptiff.sas7bdat (Imel and Probst, 2012)
- CCM Final Housing Unit Census Address List file fhum_etff.sas7bdat (Imel and Probst, 2012)

All of the above files were reduced to the collection blocks in the CCM sample, for all 50 U.S. states and the District of Columbia, excluding Puerto Rico.

3.4.1 Address Canvassing Add Actions Data

To define the relevant collection blocks in sample, the CCM Sample Design File (SDF version 103) was merged with the collection block file to subset on collection blocks that fell into the CCM sample clusters. This file was then merged by state, county, and collection block with the Combo file (Ward, 2011). The Combo file contains the AC actions, after Geography processing, for each record. If a record had an AC add action, it was included in the analysis file. If a record was marked as having an add action, even if it was a move, it was not differentiated from a true add action. The analysis file was merged with the CCM IHU and FHU files using the variables CLUSTER and Master Address File Identification (MAFID) to obtain each record's final available status. If a record was removed before FHU, its IHU status was used since this is the last known status. The AC Add Analysis was conducted using the final product, the AC Add Analysis file.

3.4.2 Address Canvassing Delete/Duplicate Actions Data

To define the collection blocks in the CCM sample, the CCM SDF (SDF version 102) was merged with the CCM sample collection block file, by the variable CLUSTER, to identify which collection blocks existed in each cluster. The CCM SDF (SDF version 102) also contained the sample weight for each cluster. In order to meet budgetary and timeline constraints, the DSSD took a sub-sample of the already-reduced CCM sample, hereafter called this evaluation of Address Frame Accuracy and Quality (AFAQ) AC sub-sample. The original CCM sample was divided into 20 (before CCM reduction) replicates or random groups. Each of these random groups was a nationally representative sub-sample. Five random groups were selected for the AFAQ AC sub-sample. This sample was only 25 percent of the original CCM sample (originally 20 random groups), and less than 50 percent of the reduced CCM sample. Such a small sample yielded unreliable and imprecise estimates of some subcategories of the AC Delete Error Rate

because the variance of the estimates was so large. As a result, some tables in the results section omit the weighted counts.

After merging the clusters with the collection blocks, the DSSD merged the file with the GQV state level files, which contained the post processed (processing performed by Geography Division) AC actions on state, county, and collection block. The GQV files are a similar vintage to the initial UCM files. Only AC records indicating a delete or duplicate action that fell into the AFAQ AC sub-sample were included, for a total of 30,476 records. After NPC clerical matching and field followup had been completed, the delete file was merged with the CCM IHU IL and CCM census address list by the variables CCM Map Spot Number and MAFID, respectively.

After clerical matching, field operations, and after followup coding, the excel files for each were all combined to create the AC Delete/Duplicate Analysis file.

3.5 Methodology for Address Canvassing Adds Actions

This section describes the categories of CCM match codes that this evaluation defines as correct adds and units added-in-error.

3.5.1 Correctly Added Address Canvassing Units

In general, a unit is a correct add if it represents a valid, non-duplicate housing unit on Census Day that is geocoded to the cluster or surrounding ring of blocks. This evaluation also gives census the benefit of the doubt and defines housing units with unresolved statuses or incomplete information on their follow-up forms as correct adds. Further, if a duplicate status has not been confirmed by field work, this evaluation considers it a correct add. For more information, including the actual CCM match codes, please refer to the Appendix.

3.5.2 Address Canvassing Units Added-in-error

In this section, the CCM match codes that indicate a unit added-in-error are defined. If an AC Add had a CCM match code that indicates it was not a housing unit on Census Day, that it was geocoded outside the ring of blocks that surround the block cluster, or field work confirmed it was a duplicate, these records were added-in-error. For more information, including specific CCM match codes, please refer to the Appendix.

3.6 Methodology for Address Canvassing Delete/Duplicate Actions

The methodology for AC Delete/Duplicate analysis consisted of six major steps:

1. Drawing a sub-sample of CCM's reduced block clusters (the AFAQ AC sub-sample)
2. Creation of file with delete/duplicate actions in CCM sample
3. NPC clerical matching of deleted records to the CCM Independent List
4. Field follow-up of matched delete records
5. After field follow-up coding by NPC
6. Weighting and estimation adjustment for this evaluation's reduction

3.6.1 AFAQ Address Canvassing Sub-sample

A sub-sample of the already-reduced CCM sample was drawn to meet budgetary and timeline constraints, as discussed in section 3.4.2 the Delete/Duplicate analysis AFAQ AC subsample.

3.6.2 File of Address Canvassing Delete/Duplicate Actions in the Census Coverage Measurement Sample

The AC records with delete/duplicate actions were subset on the CCM collection block clusters in the AFAQ AC subsample. NPC matched these records to the CCM IHU IL and sent a few cases for field work to obtain a resolution. After the field followup, NPC did the AFU coding. For a complete description of the files used in the creation of the Delete/Duplicate Analysis file see section 3.4.

3.6.3 Clerical Matching

Records deleted from AC were not in the census address list to which CCM matched their IL, so the CCM could not match their IL to these AC records. For this evaluation, NPC conducted a special clerical matching operation to attempt to match these records to the CCM IL. The goal of the operation was to search for the 30,476 AC delete/duplicate records in the CCM IL. Analysts used the HUMaRCS software, created for CCM, as a reference. The delete/duplicate records were not entered into HUMaRCS, so the final matching results were keyed into an Excel spreadsheet rather than the system.

The NPC attempted to match AC Delete/Duplicates to both linked and unlinked CCM records. The CCM status of the matched records was then used to determine if the action was an erroneous delete. See Johnson 2010 for more information on clerical matching. At the time of clerical matching, only the CCM IHU status was available.

The NPC matched 10,747 of the AC Deletes to a valid or invalid CCM address. It is important to note that when a delete matched to a CCM address, it did not automatically mean it was deleted-in-error. The remaining unmatched 19,729 AC Deletes are assumed to not exist or not contain enough information to be matched. Both instances, nonexistent units and unresolved cases, were correct deletes.

3.6.4 Field Followup

Of the 10,747 cases that matched to CCM IL records, only 144 cases required additional field followup. The majority of the matched records did not fall into one of the following categories of cases that required field followup:

- Possible matches to unlinked valid CCM records
- Matched to unlinked CCM records with unresolved Census Day, statuses
- The CCM record and AC Delete were in different blocks within the cluster and CCM did not send the record for follow-up

After the NPC completed matching, they keyed and sent an Excel spreadsheet with the 30,476 cases to the DSSD with the new match codes. The DSSD printed the field followup forms and prepared packets that included maps of the clusters and sent them to the Regional Census Centers (RCCs). Only field staff who had worked on CCM operations were hired. They were retrained using draft procedures developed for the CCM FHU Followup to locate this evaluation's follow-up cases. This evaluation's field follow-up attempted to determine 1) if the addresses represented the same unit, 2) if the unit was a valid housing unit on Census Day, and 3) in which block the record existed.

3.6.5 After Field Followup Coding

The AFU coding determined if the AC Delete matched the CCM record and Census Day status. The NPC coded and keyed the information from field into an Excel spreadsheet. NPC analysts recorded a final match code that indicated:

- If the AC Delete truly matched the CCM record
- The Census Day status of matches
- If CCM or census had the record in the correct block

After AFU coding was complete, the NPC analysts further researched units deleted-in-error for possible reasons for their deletion. The goal of this research was to ascertain characteristics of records that were more likely to be deleted-in-error. Some examples of these categories included seasonal housing, address changes, or missed units in a multi-unit. For more information on after follow-up coding see Kephart (2012a) and for delete-reason coding see Kephart (2012b).

3.6.6 Address Canvassing Delete Weighting and Estimation

If AC deleted a unit in error, and a census operation later added it, or CCM FHU followup found it was invalid, then the delete did not ultimately result in under coverage. One major limitation of the AC deleted-in-error rate was that the NPC only matched AC delete records to the CCM IHU IL and not the FHU IL, so these late adds were not included. While a FHU status was available for a tiny subset of records, a weighted count did not yield a reliable estimate. These estimates are not presented in this report.

This evaluation did not differentiate duplicate records from true deletes. Both actions typically result in a record not being included in the UCM file, the file to which CCM matches the CCM IHU IL. The units deleted-in-error rate, and all associated tables are an aggregation of both delete and duplicate actions. Records that had a CCM status indicating a group quarter (GQ, ZQ, Q3, Q2, and MQ) were excluded from the AC Delete/Duplicate analysis.

When the DSSD further reduced the CCM sample for the AC Delete analysis, the weights had to be adjusted. This resulted in larger variances and made it difficult to produce meaningful subcategories for the deleted-in-error rate. The sub-sampling took place across all block cluster sampling strata at different rates. Every stratum was multiplied by a factor equal to the number of random groups in the stratum divided by five. The table below shows the factor applied to each block cluster weight based on its stratum's reduction rate.

Table 3. 2010 CPEX Address Frame Accuracy and Quality: Address Canvassing Delete Cluster Weight Factor

CCM Strata Random Group Reduction	Adjustment Factor to Multiply Cluster Weight
Retained 20 Random Groups	20/5
Retained 18 Random Groups	18/5
Retained 15 Random Groups	15/5
Retained 11 Random Groups	11/5

Source: Konicki, 2010a.

3.6.7 Correctly Deleted Address Canvassing Units

In this section, a correct delete action is defined in terms of its CCM match code. When the NPC could not match an AC Delete to any CCM record, it was assumed to not exist or have insufficient information to be matched. This evaluation gave the census the benefit of the doubt and assumed these records were all correct delete actions. In addition, if this evaluation matched an AC Delete to a CCM address with a code indicating it was not a housing unit, uninhabitable, unresolved, was a duplicate, or it was geocoded outside of the collection block, the action was assumed to be a correct delete. If an AC Delete matched to another census record (with a different MAFID) that CCM classified as a valid housing unit, this evaluation considered this a correct delete. Also, if multiple AC Deletes matched the same valid housing unit, that CCM determined was missing-from-census, all but one of the records was considered correctly deleted.

For more information, including specific CCM match codes, please refer to the Appendix.

3.6.8 Address Canvassing Units Deleted-in-error

This section defines the CCM match codes that indicate a unit was deleted-in-error. When an AC Delete matched to a valid, non-duplicate, unlinked housing unit, this unit was deleted-in-error. This evaluation assumes if AC had not deleted the record, it would have remained in the UCM and been matched by CCM.

A small subset of AC Deletes were later added back before the creation of the UCM in January 2010. These Deletes could have been linked to CCM's IL and resulted in a linked CCM census pair, or matched a census record that CCM failed to list but confirmed was a valid housing unit. If the MAFID of the deleted or duplicated record was the same MAFID of a record on the UCM, this evaluation considered this a unit deleted-in-error. This evaluation was attempting to measure the error rate of AC Adds and Deletes. If a Lister took an erroneous action on a record, it is still an error by the 2010 Census AC even if the mistake was corrected for a subset of records during subsequent census processing and operations. If an AC Delete matched a record that was missing-from-census that was confirmed to be a valid housing unit by CCM or this

evaluation's field work, then the record was deleted-in-error. If multiple AC Deletes matched to the same valid missing-from-census housing unit record, then only one record was deleted-in-error.

For more information, including specific CCM match codes, please refer to the Appendix.

3.7 Data for Census Unedited File Adds and Deletes/Duplicates

The DSSD combined data from the following files to analyze the CUF Adds and CUF Deletes/Duplicates:

- 2010 CUF Post Capture Processing Operation File (CUFOP) (U.S. Census Bureau, 2010),
- 2010 CUF Post Capture Processing Address File (CUFAD) (U.S. Census Bureau, 2010),
- 2010 Final Tabulation MAF Operation File (MAFOP) (Zhang, 2010),
- 2010 Final Tabulation MAF Extract File (MAFX) (Zhang, 2010),
- 2010 Enumeration Extract from the 2010 Address Frame Combination (AFCOMBO) File (Ward, 2011),
- CCM Sample Design File, Version 3 (Konicki, 2011) SDF.103,
- DSSD Data File (Konicki, 2010a) SBCDATA.101,
- CCM Final Housing Unit Matching Person Interview Address Output (FHUM_PTFF) (Imel and Probst, 2012) ,
- CCM Final Housing Unit Matching Census Address Output (FHUM_ETFF) (Imel and Probst, 2012), and
- the CCM Initial Housing Unit Matching Census Address Output (HUM_CENADDFULL) (Imel and Probst, 2012).

The remainder of section 3.7 describes how the DSSD combined these data files to create a CUF Adds Analysis File and a CUF Deletes/Duplicates Analysis File.

The DSSD merged the CUFOP and CUFAD files by collection state and MAFID. The two files had the same number of records. These files contained all living quarters that were valid and enumerated in the 2010 Census plus records that census operations deleted (after creation of the January 2010 UCM). While the CUF included vacant housing units, it excluded vacant GQs. The CUF had a total of 136,757,565 (stateside) address records with approximately 131,704,730 as occupied or vacant housing units, 165,481 as occupied GQs, and 4,887,354 as invalid or deleted records.

After merging the CUFOP and CUFAD files by collection state and MAFID, the DSSD then merged the MAFOP and MAFX files by tabulation state and MAFID. The MAFOP had a record for every time a MAFID had an action in an operation/MAF source, so a MAFID may appear multiple times on the file. In all, the MAFOP had 214,721,813 records including duplicates that were in a 2010 Census operation. The MAFX had one MAFID record for all records on the MAF, both valid and invalid. Its record count was 189,971,689 total records. However, some of the MAFX records did not have a matching MAFID on the MAFOP. Most of these appeared to be duplicate records with "surviving" MAFIDs, and other records appeared to be ungeocoded.

Thus, these records did not match and were not valid records for the 2010 Census, and are outside the scope of this evaluation.

To determine the adds and deletes/duplicates, the DSSD compared the CUF to the Enumeration Universe (EU) from the AFCOMBO file as described in sections 3.8 and 3.9. There was a difference in timing between the creation of the EU Extract and the UCM used by CCM as the universe for matching in the IHU. The EU Extract was created in mid-December 2009 (Imel, 2010) and the UCM for CCM IHU matching was created in January 2010 (Cantu, 2009). As a result, there may be a small difference between the universe of adds and deletes/duplicates as defined by CCM and by this evaluation.

Before matching the CUF Adds File and the CUF Deletes/Duplicates file to the CCM IHU and FHU results, the DSSD matched the CUF collection block in these files to the CCM SDF (Version 3) to pick up the cluster weight variable for records where the CUF collection block was in a CCM sample cluster. This match included a merge with the Sample Block Cluster Data File, because the SDF did not include the CCM cluster. Note that since the CUF collection geography was not updated for records that moved state or block, the analysis may have missed matches to CCM records for MAFIDs (especially for CUF Adds) that the census operations moved to a different block. The analysis included only the records that matched CCM units.

The DSSD then matched the CUF Adds file and the CUF Deletes/Duplicates file to CCM results from IHU and FHU. The analysis results refer to these files as the CUF Adds Analysis Files and the CUF Deletes/Duplicates Analysis file, respectively. The DSSD coded matching records as correct or in error as described in the Appendix.

3.8 Methodology for Census Unedited File Adds

To identify CUF New Adds, as opposed to CUF Reinstated Adds, the DSSD match-merged the EU Extract from the Combo File[8] (Ward, 2011) to the CUF by MAFID. MAFIDs that appeared on the CUF as valid/enumerated records (PP_CUF='1'), but were not on the Enumeration Extract were considered to be New Adds for this evaluation. The CUF Reinstated Adds were valid/enumerated CUF records that existed on the Enumeration Extract as ineligible according to the filter criteria (see Imel, 2010 for details on the filter criteria). As a result, these records did not continue on to other census operations. Since they were valid/enumerated on the CUF, one (or more) of the census operations added the addresses and the addresses were successfully matched back to existing MAFIDs. The combination of the CUF New Adds and the CUF Reinstated Adds comprise the CUF Total Adds.

The DSSD matched both the CUF New Adds and the CUF Reinstated Adds to the CCM FHU results to determine whether the census correctly added the units or added the units in error. Units that FHU classified as correct enumerations or that had an unresolved status were correct adds. Units that FHU classified as erroneous enumerations were units added-in-error. Details of

[8] The 2010 Address Frame Combination File (Combo File) combined eight MAF--related data files, including the 2010 EU Extract.

the FHU match codes for correct enumerations, unresolved units, and erroneous enumerations are in the Appendix.

As mentioned in section 3.7, there was a difference in timing between the creation of the EU Extract and the UCM used by CCM as the universe for matching in the IHU. This difference in timing means that CCM processed, in IHU, some of the cases considered as New Adds by this evaluation. For the CUF Reinstated Adds, the addresses did not get passed to the UCM, so they were not included in the IHU matching. However, in FHU, CCM processed the records added by census operations after creation of the UCM. Some of the CUF Adds matched addresses that did not have a FHU final match status for these cases.

3.8.1 Correctly Added Census Unedited File Units

CUF Adds matching addresses that CCM assigned with either correct enumeration match codes or match codes of unresolved were correctly added units. Correct enumerations were addresses that CCM confirmed as valid housing units that existed on Census Day in the cluster or surrounding ring of blocks, while the unresolved addresses were cases where CCM could not confirm status of a unit with certainty. For this evaluation census is given the benefit of the doubt and unresolved cases are assumed to be correct adds. If CCM suspected a record was a duplicate, but it was not confirmed in the field, this evaluation assumed these records were not duplicates and were correct adds.

For further information about CCM match codes please refer to the Appendix.

3.8.2 Census Unedited File Units Added-in-error

In this evaluation analysis, units added-in-error were CUF Adds matching units identified in CCM as erroneous enumerations. The erroneous enumerations refer to units added to the census inventory that CCM operations verified as not existing on April 1, 2010 in the sample block cluster or the surrounding blocks. If CCM verified a record was a duplicate or not a housing unit, this evaluation considered it to be added-in-error. In addition, if a record was geocoded outside of the surrounding ring of blocks, it was considered to be added-in-error.

For further information about CCM match codes please refer to the Appendix .

3.9 Methodology for Census Unedited File Deletes/Duplicates

To identify the CUF Deletes, the DSSD matched the CUF to the EU Extract from the Combo File (Ward, 2011) on MAFID. The CUF Deletes were MAFIDs that were eligible on the EU Extract according to the filter criteria (see Imel, 2010 for details of the filter criteria) and the CUF identified the units as deletes (the FINAL_STATUS variable equaled 3).

One limitation of the analysis was the difficulty in identifying duplicates based on information from the census files (i.e., the CUF and Tabulation MAF). Generally, the Surviving MAFID and the Unit Status variables identify duplicates. However, census operations did not always update these variables. For example, in AC, when a unit was labeled a duplicate, the software on the

Handheld Computer did not link duplicates, so the Surviving MAFID was not populated. The Geography Division did attempt to link records and identified duplicates in post-census processing, but a file with these results was not available at the time of analysis.

However, the DSSD was able to identify some duplicate records. In comparing the CUF and EU Extract, the DSSD created a file of eligible EU Extract records where the MAFID did not appear on the CUF. This file had a total of 167,868 records. Since vacant GQs did not appear on the CUF, the DSSD expected the file to include these cases. When researching the addresses using the MAF Browser, the DSSD discovered vacant GQs as well the following types of units:

- units originally identified as GQs that were nonexistent, nonresidential, and duplicates,
- transient locations/units that were nonexistent, and
- housing units that were duplicates.

The DSSD appended these records to the CUF Deletes file to include them in the analysis. Records not matching CCM records did not end up in the analysis.

3.9.1 Correctly Deleted Census Unedited File Units

The DSSD classified a unit as correctly deleted if CCM coded the unit as an erroneous enumeration, a duplicate, or with an unresolved enumeration status. An erroneous enumeration code indicated a unit that did not exist, existed outside of the cluster and was misgeocoded, or was not a housing unit on Census Day.

For more information about CCM match codes please refer to the Appendix.

3.9.2 Census Unedited File Units Deleted-in-error

The DSSD classified units as deleted-in-error if CCM operations confirmed the housing units existed on Census Day in the cluster (i.e., correct enumerations) and they were not linked to a duplicate.

For more information about specific CCM match codes refer to the Appendix.

3.10 General Weighting and Variance Estimation

Since this evaluation is based on the CCM area sample, CCM weights were used for the estimates in this report. In order to weight the estimates, the DSSD applied one of two CCM weights: 1) a cluster-level weight called WEIGHT2 or 2) a housing unit-level weight called WEIGHTE. These weights were adjusted for the CCM sample reduction and small block cluster sub-sampling; however, WEIGHT2 does not account for the CCM E sample, only WEIGHTE does. For more information on the CCM sample design, see Konicki (2010b).

In this report, if an un-weighted count was less than 250 units, weighted estimates were not presented for the cell. For documentation purposes, only the un-weighted counts of categories with fewer than 250 cases are presented.

FHU Estimates Compared to IHU

In order to estimate the FHU status of Census operations' delete and adds, a different weight adjusted for the CCM sub-samples was needed. Several sub-samples were taken from various address lists within the CCM clusters, including the P sample[9], E sample[10], and Person Interview[11] samples. The DSSD chose to use the E sample because it was drawn from the CUF. The CUF Add and Delete/Duplicate Analysis files were created from the CUF as well, so conceptually it was closest to the E sample. For more information on the specific files used in the creation of the CUF analysis file see section 3.7. In this report, when the FHU status is presented, it is the E sample, with the E sample weights. The weights used for the estimates were as follows:

- AC Delete/Duplicate estimates used the IHU CCM status and the cluster-level WEIGHT2.
- AC Add and Geocoding Error estimates both used the IHU/FHU hybrid CCM status and the cluster-level WEIGHT2.
- CUF Add and Delete/Duplicates estimates used WEIGHT2 for their hybrid IHU/FHU status and the housing unit-level WEIGHTE for FHU only statuses.

Table 4 compares the weighted estimates to the actual counts for total actions from the AC Operational Assessment report (AL OIT, 2011) and CUF. Puerto Rico is excluded from all estimates and counts. The 90 percent confidence intervals contain the true counts for all estimates in the table.

[9] The P sample is a CCM sub-sample drawn from their Independent List.

[10] The E sample is a subsample of the CUF that is meant to geographically overlap with the P sample as much as possible.

[11] The Person Interview sample is drawn by CCM from both the IL and Census address list.

Table 4. 2010 CPEX Address Frame Accuracy and Quality: Weighted Housing Unit Estimates and True Observed Counts

Count (Weight)	Weighted Count	90% Confidence Interval Lower Limit	90% Confidence Interval Upper Limit	True Observed Count
AC Adds(WEIGHT2)	10,585,463[1]	8,613,488[1]	12,557,439[1]	10,300,593[2]
AC Delete/Duplicates (WEIGHT2*AFAQ Reduction Factor)	18,445,131[3]	15,177,037[3]	21,713,225[3]	19,498,219[2]
Total Valid Records in US (WEIGHT2)	132,538,124[4]	110,475,270[4]	154,600,978	131,704,730[5]
CUF Adds Matching FHU Only (WEIGHTE)	3,601,110[6]	2,706,629[6]	4,495,591[6]	3,338,775[7]
CUF Delete/Duplicates Matching IHU & FHU (WEIGHT2)	4,877,483[8]	4,029,058[8]	5,725,908[8]	4,887,354[9]
CUF Delete/Duplicates Matching FHU Only (WEIGHTE)	4,850,528[8]	3,945,913[8]	5,755,143[8]	4,887,354[9]

[1]AC Add Analysis file.
[2]AC Assessment.
[3]AC Delete/Duplicate Analysis File.
[4]Geocoding Analysis File.
[5]Housing Characteristic Report 2010.
[6]CUF Add Analysis files.
[7]Match of Enumeration Extract and CUF.
[8]CUF Delete/Duplicate Analysis Files.
[9]CUF.
Puerto Rico is excluded from all statistics in this table.

Variance

In this evaluation, to account for the complex sample design's effect on the variance estimates, the authors used a Jackknife variance technique (Miller, 1974). Replicate weights are simply the weights for an estimate with one sub group of the sample, or replicate, deleted and the other weights adjusted. Replicate weights were used to produce multiple estimates from the same sample, as if different samples were drawn from the same population. The variation among these estimates was used to calculate standard errors and standard deviations for this evaluation.

To produce replicate weights, the DSSD:

1. Sorted the CCM SDF on the systematic sample order variable, ORIGORDER.
2. Assigned replicate groups by sequentially numbering the sorted block clusters 1 to 100.
3. Calculated 100 replicate weights for each sample unit by setting the replicate weight for the deleted replicate group to zero and adjusting the other replicate weights appropriately.
4. Calculated a delete-a-group jackknife variance estimate using the 100 replicate weights.

All variances were calculated in SAS 9.2 for Linux using the above replicate weights with a Jackknife estimation procedure (proc surveyfreq).

3.11 Methodology for Geocoding Error

The following are the steps in this evaluation's geocoding error methodology:

1. Conducting sample selection and creating geocoding analysis files
2. Defining the extended search area
3. Processing
4. Conducting clerical matching, fieldwork, and after-followup coding

3.11.1 Sample Selection and Geocoding Analysis Files

The DSSD used the CCM reduced sample of 6,148 sample block clusters excluding Puerto Rico. If CCM declared that a valid housing unit or GQ existed in the sample and was missing from the census, for this evaluation the address was flagged, and a search was conducted in a larger search area on the MAF.

The files used for this portion of this evaluation are as follows:

- CCM Sample Design File, V3 (Konicki, 2011) SDF.103
- FINCOLMAFX10_**.sas7bdat Final 2010 Collection MAF (U.S. Census Bureau, 2010a)
- CCM Final Housing Unit Matching Person Interview Address Output (FHUM_PTFF) (Imel and Probst, 2012) ,
- CCM Final Housing Unit Matching Census Address Output (FHUM_ETFF) (Imel and Probst, 2012), and
- CCM Initial Housing Unit Matching Census Address Output (HUM_CENADDFULL) (Imel and Probst, 2012).
- The buffer block files (Johnson, 2011a)
- 2010 CUF Post Capture Processing Operation File (CUF) (U.S. Census Bureau, 2010a).

The 2010 Final Collection MAF had to be supplemented with late census adds from the CUF. These files were used to populate the extended search area for this evaluation.

3.11.2 The Extended Search Area

CCM searched within the block cluster and one surrounding ring of blocks for census housing units. This evaluation searched in rings of up to five kilometers[12] (km) around the block clusters.

[12] The search buffer sizes were arbitrarily chosen by DSSD.

Since collection blocks and block clusters do not come in a standard geographic/spatial shape, the DSSD requested that the Geography Division create "buffer rings" of blocks that traced around the shape of the block clusters. In order to produce estimates of how far from the original block geocoding error typically occurred, the DSSD requested buffer rings of one km, three km, and five km, beyond the edge of the cluster. If a surrounding block was partially or wholly contained within a particular buffer zone, the entire block was included in that buffer ring. For more information on the buffer specification see Johnson (2011a).

3.11.3 Stages of Processing

The DSSD used the 2010 Final Collection MAF and the CUF to create a master address inventory to load into the extended search area for clerical matching. Address records that were included in the extended search area had to meet the following criteria:

- All records had to have a collection block (ungeocoded records were excluded).
- All records from the Collection MAF had to contain a house number, street name, or location description.
- All records that were late adds and only on the CUF could not have a final census status indicating a delete or invalid record (final status=3).

Subsampling, for the P sample, E sample, and Person nterview Samples, occurred within large block clusters between IHU and FHU. This resulted in the potential sample size of valid and missing-from-census units (CCM match codes CI, UI, and ZQ) shrinking from about 15,000 cases to about 4,600 cases. Misgeocoded cases represented such a small percentage of missing-from-census cases in Census 2000 that it was assumed a larger sample than 4,600 cases was needed. This evaluation chose to include the full sample of 16,000 missing-from-census cases as of IHU to produce a reliable estimate of geocoding error. However, records that had a valid CCM code in IHU and then an invalid code in FHU were mistakenly flagged for review. These cases were not included in the geocoding error rate.

Gunnison Consulting Group Inc. (Gunnison), was contracted to modify the CCM HUMaRCS software. The modified software was named the Housing Unit Evaluation Matching Review and Coding Software (HUEMaRCS). Gunnison performed the following tasks during pre-processing:

- Loaded all valid CCM IHU/FHU linked and unlinked IL/ census records, while excluding duplicates and other invalid CCM codes.
- Loaded all additional census addresses from the MAF/CUF file mentioned above, that were geocoded to one of the blocks in the extended search area.
- Flagged IHU/FHU cases that were considered valid and missing from census (CCM codes CI, UI, and ZQ) for review by NPC analysts.
- Flagged linked CCM/census pairs that had discrepant blocks within the cluster for block review by NPC analysts. In a block cluster with multiple blocks, census and CCM could

specify different blocks for the same record in the cluster. If CCM placed the record in the correct block, this evaluation classified it as geocoding error.
- Coded linked CCM/census pairs that had the census record in the surrounding ring as geocoding error.
- If the records were in the same block, addresses were coded as a match to the block or "MB" and sent on to AFU.

3.11.4 Matching and Field work

The NPC conducted all clerical matching and AFU coding for the geocoding error rate. They used the modified HUEMaRCS software to search by cluster for missing records in the extended search area.

For linked records, if census and CCM recorded the same address in discrepant blocks within a cluster, an NPC analyst attempted to ascertain in which block the record existed. Some of these cases were sent to field to resolve the discrepancy. If the census record was correct, the case was coded to indicate it was not a census geocoding error. If CCM was correct, the record was coded to represent a geocoding error.

There were 478,555 unweighted address records in the Geocoding Analysis File. This universe consisted of:
- The processing/clerical workload of 457,141 cases.
    - Of the missing-from-census cases, 5,001 were found during clerical matching. If a case was not found, it was excluded from the analysis file. An additional 452,382 linked CCM and census pairs had to be checked for discrepant geocoding. The majority of these cases were coded automatically by HUEMaRCS during processing, and only a small subset required clerical coding by an analyst.
        - Field followup was required for 1,873 cases, either to confirm a case that was found was a true match, or to resolve a discrepancy in the block code between CCM and census.
- A small number of valid CCM records, 21,414, which were mistakenly excluded from processing and clerical work.

Field work was conducted out of the Regional Offices (ROs) instead of the RCCs, because the RCCs had begun to close. Headquarters' Technology Management Office (TMO) staff modified the CCM Operations Control System and deployed it to the ROs, when it was previously only in the RCCs. Field staff with CCM experience was recruited, and they were retrained on the modified CCM FHU field training materials.

3.11.5 Geocoding Error Rate

Two types of geocoding errors exist:
- *Geocoding errors of exclusion* – a unit that existed in a sample block but was geocoded on the MAF outside of that sample block.

- *Geocoding errors of inclusion* – a unit that existed outside of the sample block but it was geocoded on the MAF inside of the sample block. This is the geocoding error estimate produced by CCM in Keller and Fox 2012.

In this evaluation, the DSSD only attempted to measure and report on *geocoding errors of exclusion*.

For the CCM geocoding error rate, cases found within discrepant blocks in the cluster were not included. In this evaluation, if the census block was incorrect it was considered geocoding error whether the block was within the cluster, the surrounding ring, or the buffer search area. The geocoding error rate was composed of the number of erroneously geocoded addresses on the MAF divided by the number of valid census units that were confirmed to exist in the cluster by CCM. All cases in the numerator were included in the denominator.

3.11.6 AFAQ Geocoding Error Match Codes

This evaluation used the following match codes to indicate the geocoding status of an address record:

Geocoding error within the CCM search area or MA - The unit existed in the block that the CCM IL recorded it, but census had mistakenly geocoded the unit to a discrepant block in the cluster or surrounding ring of blocks. This evaluation considered this a geocoding error and included it in the numerator for the geocoding error rate.

Geocoding is correct or MB - Both census and CCM agreed on the block the unit existed in and it was within the cluster.

Geocoding is correct or MC - Census and CCM had the unit in discrepant blocks within the cluster and census was correct.

Geocoding error outside the CCM search area or MT - The unit existed within the cluster but census had the unit in a block in the extended buffer search area.

Unresolved potential geocoding error within the CCM search area or DU - The CCM IL and census placed the record in discrepant blocks in the cluster or surrounding ring and the correct block could not be determined.

Unresolved potential geocoding error outside the CCM search area or TU - The CCM IL record was linked to a census address outside the cluster in the extended search area and field was unable to determine if the case was truly an MT.

Unresolved within or beyond CCM search area or BU - The CCM IL record was a possible match to a census record within the cluster or buffer search area, but it could not be determined if the addresses referred to the same unit.

An unlinked valid and missing from census CCM record or NI - a) The CCM IL record was not linked to a census record during this evaluation or b) the Field Division confirmed a possible match did not refer to the same address.

An unlinked valid census record or NE - a) The census record was not linked to a CCM IL record during this evaluation or b) the Field Division confirmed a possible match did not refer to the same address.

3.11.7 Geocoding Error Rate Formula

CCM did not consider a case to be geocoding error if it was found within the cluster or surrounding ring of blocks. CCM only estimated geocoding error for records that were incorrectly geocoded to the block cluster or surrounding ring when they existed outside the search area. In contrast, this evaluation calculated an error rate that included:

- Discrepant block cases - In a multi-block cluster, CCM had the unit in block A, but census had it in block B. It was then field-confirmed by CCM, or this evaluation, which was the correct block. If CCM was correct, it was included in the numerator as geocoding error.
- Surrounding block cases - CCM linked the unit to a census case that existed in the surrounding blocks to the CCM cluster.
- Buffer ring block cases - This evaluation found the unit on the MAF in the extended search area outside the surrounding blocks to the cluster. It was then field confirmed that this census address did not exist in the extended search area.

The formula includes the following codes from section 3.11.6: MA+MT/MA+MT+MB

Unresolved Cases

To handle unresolved cases, the DSSD calculated three geocoding error rates:

1) a best-case scenario where all unresolved cases did not represent geocoding error,
2) a worst-case scenario where all unresolved cases did represent geocoding error, and
3) a scenario that excluded all unresolved cases. When calculated and rounded, all three error rates resulted in the same value of 1.5 percent. The authors decided to use the best-case scenario and classify all unresolved cases as not representing geocoding error.

3.12 Assumptions

The assumptions for this evaluation are as follows:

- The CCM results, supplemented by clerical matching and fieldwork in this evaluation, represent ground truth.

- The sample and weighting designs give proportions that are similar to the population proportions, even if the weighted counts are different from the population counts.

4 Limitations

4.1 Scope

Puerto Rico, and Remote Alaska were out-of-scope for this evaluation. Remote Alaska was excluded from CCM. Since this evaluation used data from CCM, the DSSD did not have information for these areas. Puerto Rico was excluded due to the cost and logistics of conducting two field operations in that area, in addition to the time and resources needed to translate training materials and forms into Spanish.

4.2 Sampling Error

The results from CCM used in this evaluation are based on a sample survey. Therefore, the estimates are subject to sampling error, and may deviate from the population values. This evaluation reports weighted estimates that take into account the complex design of CCM in order to reflect variations due to sampling; however, they do not account for nonsampling errors. Hence, the standard errors provide an indication of the minimum amount of possible error present in the estimates.

The CCM survey was not designed to measure the estimates produced in this evaluation. This evaluation applied CCM cluster- or housing unit- level weights without adjusting the weights to the population (or universe) totals. As a result, the weighted estimates tend to underestimate the population counts. However, the weighted percentages accurately reflect the population proportions. In addition, some of the categories measured in the analysis are smaller than the threshold for statistical precision. Estimates that are based on sample counts too small (250 or less) to produce statistically reliable estimates were omitted.

4.3 Nonsampling Error

Nonsampling error is a catch-all term for errors that are not a function of selecting a sample. They include error that may occur during data collection and processing survey data. Unlike sampling error, nonsampling error is difficult to quantify. Specific types of nonsampling errors that may affect this evaluation are described below.

This evaluation used a clerical operation to match addresses. Clerical matching can be subjective and is a source of nonsampling error. In addition, the clerical matchers were limited in their ability to match non-city-style address information (such as location descriptions) and to match non-city-style addresses with city-style addresses. Matching can be especially difficult for units outside of the sample block cluster. Map spots and block codes are key identifiers for units with non-city-style addresses. Thus, this evaluation may miss instances of listing error (e.g., where a lister incorrectly deleted a unit) or geocoding error for these types of addresses.

The two field operations for this evaluation were conducted one to two years after the original listings (AC and the CCM IL). Changes on the ground, specifically demolished units, may result in an underestimate of deleted or misgeocoded units. For instance, the DSSD may have sent a unit to field followup that was demolished since the original listing. If the followup staff was

unable to locate the unit and verify that it existed on April 1, 2010, this evaluation may have incorrectly considered it "correctly deleted" or "correctly geocoded."

After all field work and the AC Delete/Duplicate clerical matching operations were completed, the results had to be entered, and keying errors were introduced. There was no formal Quality Control process, so the authors verified the data. The authors corrected errors discovered from the AC Delete/Duplicate clerical matching, but other errors could potentially exist in the data.

The geocoding portion of this evaluation searched for "valid and missing-from-census" cases from IHU and FHU. Due to a programming error, cases that were "valid and missing-from-census" during IHU and were later found to be invalid during FHU were included in the clerical workload. These cases were removed from the final analysis and the geocoding error presented.

The CUF used for the CUF Add and Delete/Duplicate analyses, contains known variable errors, including an incorrect MAF source variable (which was derived from the Operation Code variable) and Non-Response Follow-up Vacancy Status variable.

4.4 2010 Census Address Canvassing Delete and Move Actions

If an AC delete/duplicate action resulted in a valid housing unit not being present on the UCM, it was considered deleted-in-error. As a result, if an AC lister deleted a unit from a block because it did not exist in that block, but it existed in the adjacent block, and AC failed to add the unit to the correct block, it was deleted-in-error.

4.5 Differences between Census 2000 and the 2010 Census

The AFAQ Evaluation repeats and expands on the Census 2000 evaluation titled, "Analysis of Deleted and Added Housing Units in Census 2000 Measured by the Accuracy and Coverage Evaluation" and the evaluation titled, "An Assessment of Addresses on the Master Address File 'Missing' in the Census or Geocoded to the Wrong Collection Block." Where possible, the DSSD followed the methodology used in those evaluations. However, differences in procedures exist between Census 2000 and the 2010 Census and between the A.C.E. and the CCM. As a result, details in the 2010 Census evaluation methodology vary from the Census 2000 evaluations, so estimates between the evaluations may not be directly comparable.

4.6 Unanswered Study Question

This report was unable to answer the following study question: **How many addresses coded as "missing" from the Census by CCM were on the MAF but excluded from the census?**

In Census 2000, a weighted count of 1.3 million housing units were erroneously excluded from Census due to the DMAF filter. However, to prevent the filter from causing a similar loss of potentially valid units, the 2010 Census AC filter was designed to allow census to validate questionable units. This evaluation was unable to include these excluded records in the matching universe for many of the same reasons why AC excluded them. For example, if they did not contain a collection block, the authors could not load them into the search area, or if a record has

incomplete address information, it will be difficult to match. As a result of excluding these records from the extended search area, the geocoding error rate may be underestimated.

4.7 Group Quarters

While CCM identified and matched GQ records, it did not classify GQs as correct enumerations or erroneous enumerations (with some exceptions). However, even when information on the enumeration status existed, the number of GQs in the CCM sample block clusters was too small for the DSSD to produce statistically accurate estimates.

4.8 Geocoding Error Search Area

For measuring the geocoding error, the DSSD/NPC searched for the "missing" Census units in a geographic area surrounding the CCM search area. However, it was possible for a "missing" census unit to exist outside of even this search area. As a result, the geocoding error measure may be underestimated.

4.9 Addresses Not Found in the MAF Block

If field staff did not locate a linked MAF address in the MAF collection block during the field followup for this evaluation, the analysis assumed the address actually existed in the cluster in which the CCM operations listed it. Presumably, the CCM operations confirmed that these units existed in the sample block cluster. However, it is possible that the unit existed elsewhere, that is, in neither the MAF nor CCM block. If that is the case, even though it was still a case of geocoding error on the MAF, it should not be part of the evaluation sample because the sample consisted of units that existed in the CCM sample block clusters. However, the evaluation field procedures did not allow for that distinction.

4.10 P Sample Nonmatches

The geocoding error operation did not limit the clerical matching to only residential MAF addresses. The procedures allowed the analysts to link P sample[13] nonmatches to any address on the MAF that appeared to be a match. In the field followup, the procedures did not ask field staff to make a determination about the residential status of the unit. Thus, the analysis assumed that the CCM operations collected enough information to determine that the address was a valid, residential unit on Census Day.

4.11 E Sample Nonmatches

E sample[14] nonmatches coded as correct enumerations in CCM operations were only verified to exist in the CCM cluster, not the specific block. This evaluation checked the matches from the CCM operations for geocoding error within the CCM cluster, but it did not check E sample

[13] The P sample is a CCM subsample drawn from their Independent List
[14] The E sample is a sample of the CUF that is meant to geographically overlap with the P sample as much as possible

correct enumerations. Thus, for the estimation, the analysis assumed the E sample enumerations were correctly geocoded to the MAF block. This may contribute to an underestimate of the geocoding error.

4.12 Census Day Status and Initial Housing Unit

Records that did not have a FHU status did not have on official Census Day Status. For some measures, this evaluation assumes their IHU status was still valid by Census Day and used it as a proxy for Census Day status. It was possible that a record's status could change between IHU and Census Day, and this evaluation would not measure this.

5 Results

5.1 Summary of Census Actions

The 2010 Census Program for Evaluations and Experiments (CPEX) operational assessments of the 2010 Census operations documented the actions taken on addresses. Table 5 below summarizes these actions as reported by the assessments. Even though the table does not include all 2010 Census operations, it provides an indication of the scope of actions on addresses in the census.

Table 5. 2010 CPEX Address Frame Accuracy and Quality: Actions for Records in the 2010 Census from 2010 Operational Assessments

Operation	Add (A)	Delete/ Duplicate (D, L, S)	Move (M)	Non-Residential (N)	Verify/ Change (C, K, V, T)	Total
Total (HU & GQs*)						
LUCA[1]……....	38,391,618	n/a	NA	33,403	n/a	n/a
Ungeocoded Resolution[2]…	NA	NA	NA	NA	251,258	251,258
INFOCOMM……...	25,242	NA	NA	NA	NA	25,242
HU Misses Count Review...	63,311	NA	NA	NA	NA	63,311
Address Canvassing[3]……..	10,776,894	20,457,786	5,450,563	1,238,260	117,244,302	155,167,805
Group Quarters Validation[4]	510,448	821,648	n/a	51,662	n/a	2,543,972
New Construction……..	332,603	NA	NA	NA	NA	332,603
Field Verification[5]……...	NA	234,540	NA	NA	221,153	455,693
NRFU/ Vacant Delete Check[6]	811,985	1,635,214	NA	443,073	n/a	n/a
Non-ID Operations[7]……	967,842	NA	NA	NA	n/a	n/a
Housing Unit						
LUCA……...	n/a	n/a	n/a	n/a	n/a	n/a
Ungeocoded Resolution …..	NA	NA	NA	NA	251,258	251,258
INFOCOMM……	25,242	NA	NA	NA	NA	25,242
HU Misses Count Review…	63,311	NA	NA	NA	NA	63,311
Address Canvassing…….	10,537,246	20,337,814	5,447,517	1,222,360	117,127,076	154,672,013
Group Quarters Validation...	491,096	n/a	n/a	n/a	n/a	n/a
New Construction……..	332,603	NA	NA	NA	NA	332,603
Field Verification……	n/a	234,540	NA	NA	221,153	455,693
NRFU*/ Vacant Delete Check	811,985	1,635,214	NA	443,073	n/a	n/a
Non-ID Operations	n/a	NA	NA	NA	n/a	n/a
Group Quarters						
LUCA……..	n/a	n/a	n/a	n/a	n/a	n/a
Address Canvassing…...	239,648	119,972	3,046	15,900	117,226	495,792
Group Quarters Validation……	19,352	n/a	NA	NA	n/a	n/a
Non-ID Operations……..	n/a	NA	NA	NA	n/a	n/a

n/a – not available, HU & GQs -Housing Units and Group Quarters, NRFU-Nonresponse Followup.
NA – Not Applicable in 2010 Operational Assessments.
[1] 2010 LUCA Assessment.
[2] 2010 NRFU Assessment. Table includes INFOCOMM, HU Misses Count Review, and New Construction.
[3] 2010 AC Assessment . Transitory Locations= Housing Units. Special Places=Group Quarters. Uninhabitables = Delete/Duplicate.
[4] 2010 GQV Assessment. Transitory Locations are counted as Housing Units.
[5] 2010 FV Assessment. Total excludes 220 units with blank actions.
[6] 2010 NRFU Assessment.
[7]2010 Non-ID Processing Assess.

Table 6 shows the original source of records flagged on the 2010 Census CUF as valid records. The DSSD assigned the original source variable based on a combination of the DSF variables, the Operation Date variable, and the MAF Source variable. However, the MAF Source variable has a limitation. The MAF Source originated from the Operation Code variable on the CUF. The intent of the variable was to identify the operation that created the action on an address. As documented in Pennington et al. (2011), analysts discovered that this was filled with invalid or incorrect codes on many occasions, because the data collected in the field was not validated. Where possible, the variable was repaired. The intent of the original source variable was to show the first operation or source to contribute the record to the MAF. Note that due to this limitation, the original source values for operations after Group Quarters Validation may not be accurate. Highlights from the table include:

- The DSF and Operations before the 2010 Census combined were the original source for more than 93 percent of the housing unit and GQ records on the 2010 Census CUF.

- Of the 2010 Census operations, the operation contributing the most records was AC, with nearly five million living quarters or almost four percent of all addresses on the Tabulation MAF.

- After AC, the next largest source of addresses was the LUCA, with 2.2 million or about 1.7 percent of the living quarters having LUCA as the original source.

Table 6. 2010 CPEX Address Frame Accuracy and Quality: Original Source of Census Unedited File Valid Records

MAF Source by Type of Record	Count	Percent
Total Valid CUF Records	**131,870,211**	**100.0**
Blank GQHUFLAG……..	1,544	<0.1
Housing Unit[1]	**131,704,607**	**99.9**
Unknown MAF Source……	31	<0.1
Delivery Sequence File……	36,079,673	27.4
Local Update of Census Addresses/LUCA Appeals……	2,224,296	1.7
Operations/MAF Sources before 2010 Census……	87,161,387	66.1
INFOCOMM……..	6,066	<0.1
Count Review - HU Misses……	15,321	<0.1
Address Canvassing……	4,977,753	3.8
Group Quarters Validation…….	304,540	0.2
Update Leave/UUE/UUL/Remote Alaska/Remote UE/UE……	366,752	0.3
New Construction…….	106,348	<0.1
Enumeration of Transitory Locations…….	92,662	<0.1
Non-Response Followup/Vacant Delete Check……...	335,381	0.3
Service Based Enumeration Phase 2 …….	18	<0.1
Non-ID Operations[2]……	34,380	<0.1
Group Quarters	**164,059**	**0.1**
Delivery Sequence File……	32,385	<0.1
Local Update of Census Addresses/LUCA Appeals……	5,130	<0.1
Operations/MAF Sources before 2010 Census……	80,947	<0.1
Count Review – GQ…….	247	<0.1
Address Canvassing……	10,383	<0.1
Group Quarters Enumeration/Advanced Visit…….	16,633	<0.1
Group Quarters Validation……..	12,803	<0.1
Service Based Enumeration Phase 2……...	5,528	<0.1
Non-ID Operations[3]……	3	<0.1

[1] Includes records with GQHUFLAG equal to blank, 0, 4, and 5. This includes Transient Locations/Units.
[2] Includes Be Counted, Telephone Questionnaire Assistance (TQA) – Fulfillment, TQA – Interview, Group Quarters Enumeration Usual Home Elsewhere (UHE), NRFU Whole Household Usual Residence Elsewhere (WHURE), and Remote Alaska UHE/Update Enumerate UHE/Remote Update Enumerate UHE.
[3] Includes Group Quarters UHE and Nonresponse Followup WHURE.
Source: 2010 Census Final Tabulation Master Address File (MAFOP and MAFX) and 2010 Census Unedited File.

5.2 Address Canvassing Adds

This section answers the research question: **By source/operation, what proportion of 2010 Census AC "Adds" were correctly added and erroneously added, according to their CCM status?**

Add Results

In the CCM clusters, 40,946 AC Adds matched to CCM records. The added-in-error rate was calculated using the last known CCM status of a record. If a record was in FHU, then that was the status used. If a record was not in FHU, then its IHU status was used as a proxy for Census Day status. After applying the cluster-level weight, WEIGHT2, to the 40,946 AC Adds, this resulted in a weighted count of 10,585,463 add actions. The results presented here include both new AC Adds and Reinstated Adds with a non 2010 Census AC original source.

A correct add was an add action that recorded the existence of a valid housing unit on Census Day. An erroneous add was when an add action recorded a unit that was misgeocoded, a duplicate, not a housing unit, or did not exist on Census Day. Group Quarters were outside the scope of CCM, so a Census Day status was not collected on these units; they were excluded from the AC Add analysis.

To truly understand and adapt listing procedures to prevent add action errors in the future, it is important to examine the error rate by region, type of error (nonexistent, geocoding error, etc.), original source of the address, type of enumeration area, problematic blocks, and address characteristics of blocks with errors.

Table 7 examines the overall added-in-error rate. Out of approximately 10.6 million add actions, 1.7 million records, or 16.4 percent, did not represent actual valid housing units that existed in the CCM search area on Census Day. These 1.7 million added-in-error records represent only 1.2 percent of all records processed in AC.

Table 7. 2010 CPEX Address Frame Accuracy and Quality: Status of Address Canvassing Added Records

Add Status	Unweighted Adds	Weighted Adds	Weighted Percent (SE[1])
Total Adds……	**40,946**	**10,585,463**	**100.0**
Correctly Added……	34,532	8,853,529	83.6 (1.3)
Added-in-Error……..	6,414	1,731,934	16.4 (1.3)

Source: AC Add Analysis File.
[1] Standard Error.

Table 8 reveals that the overall added-in-error rate is hiding some distinct differences between regions.

- Despite the fact the South contained the most added-in-error actions, the within-region added-in-error rate revealed that the South actually had the lowest added-in-error rate of all regions at 13.8 percent.

- In contrast to the South, the Midwest had the highest added-in-error rate at over 22 percent.

- The Northeast and West were more in line with the overall added-in-error rate at 18.2 percent and 17.2 percent, respectively.

- The fact that the South contained almost half of all add actions, combined with the lowest added-in-error rate lowered the national rate overall.

Table 8. 2010 CPEX Address Frame Accuracy and Quality: AC Added-in-Error Rate within Census Region

Add Status by Region	Weighted AC Added-in-Error	Weighted AC Correct Adds	Weighted Percent Within Region (SE[1])
Total AC Adds…….	**1,731,934**	**8,853,529**	**16.4 (1.3)**
Midwest……..	342,822	1,200,567	22.2 (7.4)
Northeast……	278,667	1,249,522	18.2 (3.0)
South……..	739,342	4,621,334	13.8 (1.2)
West……	371,103	1,782,106	17.2 (2.7)

Source: AC Add Analysis File.
[1] Standard Error.

In addition to the geography of where add action errors occur, it is useful to look at the various components of the error rate. Table 9 examines the reasons a unit was added-in-error. These include categories such as geocoding error, duplicates, and nonexistent units. In this evaluation, units exposed to the elements, nonexistent and/or nonresidential are aggregated into one category with no distinction for units with the potential to change to a valid status.

- The largest component of the added-in-error rate was records that did not represent housing units on Census Day. About 11.9 percent of all add actions were not actually

housing units. These 1.3 million records represented units exposed to the elements, nonexistent units, and/or nonresidential structures.

- The second largest component of the added-in-error rate, at 4.4 percent, was field confirmed duplicates of census records. If a duplicate was not sent to field followup, it was assumed to be a correct add that represented a unique HU.

- The smallest component of the added-in-error rate was units that were geocoded outside the search area. With only 17 unweighted cases, there were not enough records to produce a valid weighted count.

- For correctly added records, the largest component was linked CCM and census pairs. Almost 8 million or 75.4 percent of all add actions matched to a linked CCM case. If a record matched to a linked CCM census pair with an unresolved status, it was considered a correct add.

- The next largest component of correctly added units was unlinked valid census records at 6.4 percent of add actions. Sometimes CCM could not find a match for a census record in the IL list, but they field confirmed the unit was valid and existed on Census Day. These are considered correct adds.

- In Smith et al. (2003), added units that fell within the A.C.E. search area, the cluster of blocks, or immediate surrounding ring of blocks, were considered to be correct adds. This evaluation applies the same standard and included these 137,375 cases in the correct add rate.

Table 9. 2010 CPEX Address Frame Accuracy and Quality: Status of AC Adds in CCM Housing Unit Results

CCM Enumeration Status	Unweighted AC Adds	Weighted AC Adds	Weighted Percent (SE[1])
Total AC Adds…….	**40,946**	**10,585,463**	**100.0**
Correct Enumerations (Correctly Added by AC)……..	34,532	8,853,529	83.6 (1.3)
Matches……...	31,333	7,984,931	75.4 (1.7)
Unlinked Valid Census Records…….	2,227	673,037	6.4 (1.3)
Geocoding Error within Search Area……..	650	137,375	1.3 (0.4)
Unresolved Units……	161	*	
Unconfirmed Duplicates……	161	*	
Erroneous Enumerations (Added-in-Error by AC)……...	6,414	1,731,934	16.4 (1.3)
Not a Housing Unit……..	4,995	1,260,086	11.9 (0.8)
Confirmed Duplicates……..	1,402	463,125	4.4 (1.2)
Geocoding Error Outside Search Area……	17	*	*

[1] Standard Error.
Source: AC Add Analysis File.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table 10 examines the added-in-error rate by type of unit.

- The vast majority of added units were single or multi-unit structures, at 6 million and 2.4 million, respectively.
- Multi-units accounted for 23.7 percent of all add actions.
- Mobile homes accounted for only about 1.5 million of the 10.6 million add actions, or about 13.9 percent of add actions; however, they made up about 17.6 percent of all added-in-error actions.
- The "other" category contains 156 unweighted correct adds and 232 erroneously added units. While there are not enough cases to estimate weighted counts, this may be qualitative evidence that "other" or transient locations have a much higher added-in-error rate than any other structure type.

Table 10. 2010 CPEX Address Frame Accuracy and Quality: Status of AC Adds by Type of Structure

Type of Structure	Unweighted AC Adds	Weighted AC Adds	Weighted Percent (SE)
Total AC Adds……	**40,946**	**10,585,463**	**100.0**
Correct Enumerations…….	34,532	8,853,529	83.6 (1.3)
Single-unit……	19,232	5,264,417	49.7 (2.5)
Multi-unit……..	8,811	1,963,936	18.6 (2.2)
Mobile Home……	3,608	1,169,219	11.0 (0.8)
Other…….	156	*	*
Missing/Unknown……	2,725	396,582	3.7 (1.7)
Erroneous Enumerations…….	6,414	1,731,934	16.4 (1.3)
Single-unit……	3,011	796,100	7.5 (0.6)
Multi-unit……..	2,063	546,197	5.2 (1.3)
Mobile Home……	886	305,642	2.9 (0.3)
Other…….	232	*	*
Missing/Unknown…….	222	*	*

Source: AC Add Analysis File.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

The added-in-error rate was composed of both new adds and reinstated adds that had an original source pre-dating AC.

Table 11 separated sources into three categories: AC, the DSF, and all census operations prior to AC. An added-in-error rate was calculated for all three using the weighted number of records added-in-error over the weighted counts of records from that source.

- 2010 Census AC had the highest added-in-error rate as a source at 23.5 percent. This high error rate could be due to the fact AC Lister training encouraged Listers to add when there was any doubt, or potentially a unit could be valid by Census Day. Also, AC Adds had not been validated by a prior census operation or the DSF and therefore, are more likely to not represent valid structures.

- Records from the DSF had the lowest added-in-error rate at less than 4 percent. Many of these records were valid but were not sent to AC because they were ungeocoded.

Table 11. 2010 CPEX Address Frame Accuracy and Quality: Weighted AC Adds by Original Source

Original Source	Weighted AC Units Added-in-Error	Weighted AC Correct Adds	Weighted Percent Added-in-Error (SE[1])	Total AC Adds
Total AC Adds…….....	**1,731,934**	**8,853,529**	**16.4 (1.3)**	**10,585,463**
Address Canvassing…….	1,564,244	5,104,858	23.4 (2.0)	6,669,102
DSF……..	123,496	3,345,186	3.6 (0.9)	3,468,682
Pre-AC Operations……	44,195	403,485	9.9 (2.2)	447,679

[1] Standard Error.
Source: AC Add Analysis File.

Table 12 explores the distributions of added-in-error actions by block cluster.

- A total of 2,803 CCM clusters from a sample of 6,148 CCM clusters , contained at least one AC add action. This represents about 46 percent of all clusters in sample having at least one add action.

- Over 50 percent of clusters with an AC add action contained no added-in-error actions.

- Almost eight percent (209) of clusters with an AC add action had less than 10 percent of their add actions in error.

- Over 12 percent (343) of clusters had 10 percent to 25 percent of their add actions in error.

- Almost a third of clusters with an AC add action had 25 to 100 percent of their add actions in error.

Table 12. 2010 CPEX Address Frame Accuracy and Quality: Distribution of Address Canvassing Added-in-error Actions

Error Rate within Cluster	Unweighted Number of Clusters	Unweighted Percent of Added-in-Error Actions in Cluster
Total clusters in sample with an add action	**2,803**	**100.0**
No add actions were in error…….	1,413	50.4
1 to 10 percent of add actions were in error……..	209	7.5
10 to 25 percent of add actions were in error……	343	12.2
25 to 100 percent of add actions were in error……..	838	29.9

Source: AC Add Analysis File.

5.3 Address Canvassing Deletes/Duplicates

Deletes

This section answers the following research question: **By source/operation, what proportion of AC deleted/duplicated units were correctly deleted and erroneously deleted?** Unlike AC Adds and CUF Adds and Deletes, AC Deletes were measured with a sub-sample of the already- reduced CCM sample. Due to this further sample reduction (i.e., the AFAQ AC sub-sample), there were not enough cases to produce reliable estimates of several categories below. If an un-weighted cell contained fewer than 250 units, weighted estimates are not presented. For documentation purposes only, the un-weighted counts of categories with fewer than 250 cases are presented.

Table 13 presents the overall AC deleted-in-error rate. A unit was deleted-in-error for AC if its deletion resulted in a valid housing unit missing from the census address list to which CCM matched the IHU IL. If the address was added by a later census operation, the unit was still considered an AC unit deleted-in-error. NPC attempted to match AC delete/duplicate records to everything in the associated cluster and surrounding ring of blocks, so if AC deleted a record from one block but added it in another, it would have been matched to the CCM/census pair. If an AC delete/duplicate record did not match to any CCM record in the entire search area, it was considered a correct delete. The Census 2000 BC operation, as noted earlier, has many procedural differences from the 2010 Census AC operation, including, but not limited to: BC was paper based while AC was automated, BC was only done inside the blue line compared to AC which was nationwide, and most importantly, AC Deletes were validated within the operation, BC's were not. However, the BC operation is the closest that exists from Census 2000, and the number is included for discussion.

- Out of 18.4 million delete/duplicate actions, only 786 thousand, or 4.3 percent were deleted by AC in error. This represents 0.5 percent of all actions processed in AC.

- These 786 thousand records represented only half of one percent of the 132 million valid housing units AC found.

- About 17.7 million out of the 18.4 million deleted records represented one of the following correctly deleted categories: duplicates, invalid housing units, misgeocoded units, or units with unknown/unresolved status.

Table 13. 2010 CPEX Address Frame Accuracy and Quality: Status of AC Deletes/Duplicates

Delete/Duplicate Status	2000 Block Canvassing	2010 AC		
	Count & Percent	Unweighted Count	Weighted Count	Weighted Percent (SE[1])
Total Deletes/Duplicates……	**5,146,320**	**30,476**	**18,445,131**	**100.0**
Correctly Deleted……...	76	29,171	17,658,837	95.7 (0.6)
Deleted-in-Error……..	24	1,305	786,294	4.3 (0.6)

[1] Standard Error.
Source: AC Delete/Duplicate Analysis File, Burchman 2002.

Table 14 explores the distribution of delete actions by region.

- By region, there were not enough deleted-in-error cases to produce reliable weighted counts for any region, besides the South, at 442 thousand or 56 percent of the 786 thousand deleted-in-error cases.

- The South also contained 53 percent of all correctly deleted records, at 9.3 million cases.

- The other correct delete actions were spread evenly between the Midwest, West, and Northeast, ranging between 14.7 percent and 15.4 percent of all delete/duplicate records.

Table 14. 2010 CPEX Address Frame Accuracy and Quality: Status of AC Deletes/Duplicates Results by Region

Delete/Duplicate Status by Region	Unweighted AC Deletes/Duplicates	Weighted AC Deletes/Duplicates	Weighted Percent (SE[1])
Total AC Deletes/Duplicates…….	**30,476**	**18,445,131**	**100.0 (0.0)**
Correctly Deleted……...	29,171	17,658,837	95.7 (0.6)
Midwest……..	3,054	2,714,803	14.7 (1.2)
Northeast……	3,927	2,849,112	15.4 (1.5)
South……..	15,196	9,299,957	50.4 (2.2)
West……...	6,994	2,794,965	15.2
Deleted-in-Error…….	1,305	786,294	4.3 (0.6)
Midwest…….	108	*	*
Northeast……	115	*	*
South……..	848	497,267	12.3 (1.6)
West……	234	*	*

[1] Standard Error.
Source: AC Delete/Duplicate Analysis File.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table 15 breaks out the categories of units deleted-in-error and deleted correctly. These types, nonexistent unit, geocoding error, etc., are determined by the CCM IHU match code.

AC delete/duplicate records that did not match to any CCM records, were assumed to be correct deletes because a) they either did not exist or b) the records did not contain enough information to match to a valid address.

The largest category of deleted-in-error cases was valid unlinked CCM or census cases. A valid unlinked CCM case occurred because the record was mistakenly deleted from the census address list before matching. A valid unlinked census case was when CCM had a census record on the address list and they verified it was a valid housing unit even though CCM failed to list it. When the MAFID of the AC Delete was the same as the valid unlinked census case, this represented a unit deleted-in-error that still was sent to processing.

Typically, unresolved cases were considered to be correctly deleted. The one exception was five cases that were field confirmed by this evaluation to represent valid housing units on Census Day. As with unlinked valid census records, a valid CCM census pair was considered to be

deleted-in-error when the MAFID of the AC Delete was the same as the census record in the pair.

- Almost 68 percent of all AC Deletes (12.5 million out of 18.4 million) did not match to a CCM record.

- The single largest category of deleted-in-error cases was unlinked valid CCM or census cases at 780 thousand of the 786 thousand deleted-in-error actions.

Table 15. 2010 CPEX Address Frame Accuracy and Quality: AC Deletes by CCM Initial Housing Unit Matching Results

CCM Enumeration Status	Unweighted AC Deletes	Weighted AC Deletes	Weighted Percent (SE[1])
**Total AC Deletes/Duplicates……..…….. **	**30,476**	**18,445,131**	**100.0**
Erroneous Enumeration (Correct Delete)…….…...	29,171	17,658,837	95.7
Nonmatch…….....	19,729	12,530,344	67.9 (1.5)
Not a Housing Unit……	597	443,882	2.4 (0.6)
Geocoding Error……	2,932	1,335,441	7.2 (1.2)
Census Duplicates……...	9	*	*
Unresolved Units……	14	*	*
Linked CCM and Census Records……	5,452	3,138,209	17.0 (1.0)
Unlinked Invalid CCM or Census Record…….	438	197,286	1.1 (0.4)
Correct Enumeration (Deleted-in-Error)…….	1,305	786,294	4.3 (0.6)
Linked CCM and Census Records……..	6	*	*
Unlinked Valid CCM or Census Record……	1,294	780,075	4.2 (0.6)
Unresolved Unit (AFAQ Field Confirmed)……	5	*	*

Source: AC Delete/Duplicate Analysis File.
[1] Standard Error.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table 16 presents probable reasons, or categories, for units being deleted-in-error. These were determined by a special matching operation at NPC (Kephart, 2012a). After the matching operation had been completed, several additional CCM codes were added to the category of units

deleted-in-error. These records do not have a classification for delete reason. This table is unweighted, so it only presents qualitative evidence of reasons for deleted-in-error actions. Three categories account for almost 75 percent of deleted-in-error cases: 1) units that were added back by a later census operation or matched after IHU, 2) analyst unable to ascertain probable reason for deletion, and 3) hidden units in multi-units. The remaining 11 categories contain less than five percent of records each.

- The single largest category of units deleted-in-error was records that were added back by a later census operation, or were matched after IHU, at 47.1 percent.

- The next largest category was records that the analysts were unable to determine a reason for deletion, since the address does not appear to be an unusual structure or have any other probable reasons for deletion, at 18.0 percent.

- The third largest categorywas hidden units in a multi-unit structure, nine percent. Previous research has shown that hidden units in a small multi-unit are a significant problem for Listers and interviewers (Virgile, 2012).

Table 16. 2010 CPEX Address Frame Accuracy and Quality: AC Deletes/Duplicates Deleted-in-Error by Reason for Error

Reason for Delete Error	Unweighted AC Deletes	Unweighted Percent
Total	**1,305**	**100.0**
Unit was later linked by FHU (late link or late census add)……	613	47.1*
Unit deleted due to Lister error……	233	17.9*
Unit is in a multi-unit where the building was not missed but the unit was…	117	9.0*
Unit is seasonal housing……..	62	4.8*
Unit was a duplicate but no primary can be found……..	57	4.4*
Unit is a hidden unit that is a separate structure……..	56	4.3*
Building is a multi-unit that was completely missed……..	28	2.2*
Unit is in a trailer park and it is distinct from a hidden unit situation…….	23	1.2*
Record has a map and/or geography error……...	22	1.7*
Unit is new construction……..	18	1.4*
Move that the record was not properly added back……	17	1.3*
The same MAFID represents an address with significant changes…….	17	1.3*
Census thought was uninhabitable (boarded up, etc.), but CCM disagrees…	14	1.1*
Unit was not sent for delete reason coding……..	13	1.0*
Unit has slightly different address, such as hn, street, etc……	12	0.9*
Confusion if unit was a GQ……...	3	0.2*

*Percentages are based on unweighted counts.
Source: AC Delete/Duplicate Analysis File.

Table 17 explores the distribution of delete actions and deleted-in-error actions at the block cluster level. The reduced AFAQ AC Delete/Duplicate sub-sample contained 2,488 CCM clusters instead of the 6,148 clusters in the CCM sample.

- Out of the 2,488 clusters, 1,689 clusters contained at least one delete/duplicate action, this represents 68 percentof clusters in sample.

- Eighty percent of clusters that contained a delete/duplicate action had no units deleted-in-error.

- Almost 11 percent of clusters that contained a delete/duplicate action had less than 10 percent of their actions in error.

- About five percent of clusters contained 10 percent to 25 percent deleted-in-error actions.

- Just over four percent of clusters contained 25 percent to 100 percent deleted-in-error actions.

Table 17. 2010 CPEX Address Frame Accuracy and Quality: Distribution of Address Canvassing Deleted-in-Error Actions

Error Rate within Cluster	Unweighted Number of Clusters	Unweighted Percent of Deleted-in-Error Actions in Cluster
Total clusters in sample with a delete/duplicate action.....	**1,689**	**100.0[1]**
No delete actions were in error……...	1,347	80.0
1 to 10 percent of delete actions were in error…….	181	10.7
11 to 25 percent of delete actions were in error…….....	88	5.2
25 to 100 percent of delete actions were in error…….....	73	4.3

[1] Percent does not round to 100 due to rounding error.
Source: AC Delete/Duplicate Analysis File.

5.4 Census Unedited File Adds

This section answers the following research question:

By source/operation, what proportion of post-AC "Adds" were correctly added and erroneously added, according to their final CCM status?

This section provides results for the combined CUF New Adds and CUF Reinstated Adds (i.e., total CUF Adds). For details on either the CUF New Adds or the CUF Reinstated Adds, refer to the Appendix.

Matching the CUF to the EU Extract found a national total of 3,338,775 CUF New Adds and CUF Reinstated Adds combined. The results in this section provide estimates for the CUF Adds that match to FHU matching results, so the analysis excluded GQs and records subsampled from FHU. The overall ratio of CUF Adds to the national number of housing units was 2.5 percent, while the Census 2000 weighted count of 3,857,381 HCUF Adds provided a ratio of 3.3 percent. For this evaluation, the weighted total of CUF Adds matching to FHU results was 3,601,110.

Table 18 shows the distribution of the CUF Adds according to the FHU matching results and compares the 2010 Census to results from the Census 2000 evaluation.

- In the 2010 Census, about 79.6 percent of the CUF Adds were correctly added. This was lower than the 83.9 percent of the HCUF Adds correctly added in Census 2000. However, since standard errors were not available from the Census 2000 evaluation, the DSSD cannot determine whether this difference was statistically significant.

- In the 2010 Census, about 20.4 percent of the CUF Adds were added-in-error, compared to the 16.1 percent added-in-error in Census 2000.

- About 66.5 percent of the CUF Adds in the 2010 Census were correctly added CUF Reinstated Adds, and 33.5 percent were correctly added CUF New Adds.

- About 55.3 percent of the CUF Adds in the 2010 Census were CUF Reinstated Adds added-in-error and 44.7 percent were CUF New Adds added-in-error.

Table 18. 2010 CPEX Address Frame Accuracy and Quality: Status of Census Unedited File Adds in CCM Final Housing Unit Matching

CUF Add Status	Census 2000		2010 Census		
	Weighted Total (Percent)	Weighted Percent (SE[1])	Unweighted CUF Adds	Weighted Total CUF Adds	Weighted Percent (SE[1])
Total CUF Adds ……	**3,857,381**	**100.0 (n/a)**	**4,769**	**3,601,110**	**100.0**
Correctly Added ……	3,235,099	83.9 (n/a)	3,449	2,867,070	79.6 (2.5)
CUF New Adds ……	n/a	n/a	1,623	960,527	33.5 (n/a)
CUF Reinstated Adds ……..	n/a	n/a	1,826	1,906,543	66.5 (n/a)
Added-in-Error ……..	622,282	16.1 (n/a)	1,320	734,040	20.4 (2.5)
CUF New Adds ……	n/a	n/a	650	328,227	44.7 (n/a)
CUF Reinstated Adds ……...	n/a	n/a	670	405,812	55.3 (n/a)

n/a is not available.
[1] Standard Error.
Source: Smith et al, 2003 and CUF Adds Analysis Files.

Table 19 shows details of the CUF Adds in the FHU matching results.

Of the total CUF Adds, the FHU classified 79.6 percent of the addresses as correct enumerations and were therefore correctly added by the census. These included:

- About 33.6 percent of the correct enumerations were cases matched in FHU. For these cases, CCM linked the census address to an IL address and confirmed that the address existed as a housing unit on Census Day.

- About 34.2 percent of the correct enumerations were census nonmatches. CCM did not link these census addresses to IL addresses, but the field followup confirmed that the census address existed as a housing unit on Census Day.

- About 10.1 percent of the correct enumerations that were found within the search area. This means that the census geocoded the addresses to the CCM sample block cluster, but field followup confirmed that the addresses existed in the surrounding blocks and the census addresses existed as housing units on Census Day.

- An unweighted count of 146 of the correct enumeration records were unresolved cases at the end of FHU. The analysis assumed the census correctly added these addresses.

The FHU classified the remaining 20.4 percent of CUF Adds as erroneous enumerations. These were addresses the census added-in-error. These included:

- About 6.0 percent of CUF Adds were classified as "not housing units" on Census Day.

- About 11.9 percent of CUF Adds were duplicates of other addresses in the census inventory.

- An unweighted count of 229 of the erroneous enumeration units were geocoding error. For these units, the census geocoded the address to the sample block cluster, but the field followup confirmed that the address was located in a block beyond the surrounding blocks to the sample cluster.

Table 19. 2010 CPEX Address Frame Accuracy and Quality: Status of Census Unedited File Adds in CCM Final Housing Unit Matching Results

CCM Enumeration Status	Unweighted CUF Adds	Weighted CUF Adds	Weighted Percent (SE)
Total CUF Adds ……...	**4,769**	**3,601,110**	**100.0**
Correct Enumerations (Correctly Added by Census)……	3,449	2,867,070	79.6 (2.5)
Matches/Possible Matches ……..	1,530	1,209,874	33.6 (4.3)
Census Nonmatches ……	1,180	1,230,468	34.2 (5.1)
Geocoding Error within Search Area …….	593	362,132	10.1 (2.4)
Unresolved Units …….....	146	*	*
Erroneous Enumerations (Added-in-Error by Census)	1,320	734,040	20.4 (2.5)
Not a Housing Unit …….....	423	216,098	6.0 (0.9)
Duplicates …….....	668	427,662	11.9 (1.8)
Geocoding Error ……..	229	*	*

Source: CUF Adds Analysis Files.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table 20 shows the distribution of CUF Adds by census region.

According to the FHU matching results, the South census region had the highest percentage of correctly added units (42.6 percent) and the highest percentage of units added-in-error (10.2 percent).

Table 20. 2010 CPEX Address Frame Accuracy and Quality: Status of Census Unedited File Adds in CCM Final Housing Unit Matching Results by Census Region

CUF Add Status by Region	Unweighted CUF Adds	Weighted CUF Adds	Weighted Percent (SE[1])
Total CUF Adds……....	**4,769**	**3,601,110**	**100.0**
Correctly Added…….....	3,449	2,867,070	79.6[2] (2.5)
Midwest……..	546	420,086	11.7 (2.9)
Northeast…….	349	285,634	7.9 (2.0)
South…….	1,523	1,532,559	42.6 (6.2)
West…….	1,031	628,791	17.5 (5.8)
Added-in-Error…….....	1,320	734,040	20.4 (2.5)
Midwest……	172	*	*
Northeast……..	166	*	*
South……..	486	367,362	10.2 (1.4)
West……..	496	128,071	3.6 (0.7)

[1] Standard Error.
[2] Percent may not sum to applicable subtotals.
Source: CUF Adds Analysis Files.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

The Address Characteristic Type (ACT) code was a classification of a block to the predominant type of address in the block (e.g., city-style, rural route, noncity-style, etc.) and whether or not the address was carried in the DSF. The ACT code was assigned to blocks prior to the AC operation and does not reflect updates from any of the 2010 Census operations (Contreras et al., 2012). The ACT code had many values, which Table 21 has collapsed into three categories as follows:

1) "All city-style" includes blocks with no DSF coverage, some DSF coverage, and all DSF coverage.
2) "Mixed" includes blocks with 70-99 percent city-style addresses and some, all, or no DSF coverage.
3) "Non-city-style" includes blocks for all other values, such as rural route, nonresidential, post office box, and zero blocks (blocks with no addresses at the time of classification).

Table 21 shows that most of the CUF Adds were from city-style blocks (51.9 percent). The CUF Adds in city-style blocks had the highest rate of CUF Adds added-in-error, with 11.8 percent of all adds in error. Note, however, that the correctly added rate for the CUF Adds in non-city-style blocks may be an overestimate. As mentioned in the limitations, addresses in these blocks can be difficult to match and analysts may have coded a higher percentage of these as unresolved. As previously noted, the analysis treated unresolved cases as correctly added.

Table 21. 2010 CPEX Address Frame Accuracy and Quality: Status of Census Unedited File Adds by Address Characteristic Type from Final Housing Unit Results

Collection Block Type of Address	Unweighted CUF Adds	Weighted CUF Adds	Weighted Percent (SE[1])
Total CUF Adds…….....	**4,769**	**3,601,110**	**100.0 (0.0)**
Correctly Added…….....	3,449	2,867,070	79.6 (2.5)
City Style…….	2,061	1,869,845	51.9 (5.6)
Mixed……	850	609,362	16.9 (2.9)
Non-city Style……...	538	387,863	10.8 (4.1)
Added-in-Error…….....	1,320	734,040	20.4 (2.5)
City Style……..	744	425,177	11.8 (1.7)
Mixed……	414	227,910	6.3 (1.0)
Non-city Style……	162	*	*

[1] Standard Error.
Source: CUF Adds Analysis Files.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table 22 shows the weighted count of CUF New Adds by the Original Source variable. The original source for most of the addresses correctly added by the census was Update Leave/Urban Update Enumerate/Urban Update Leave (22.2 percent). None of the sources for the addresses added-in-error had counts large enough to provide statistically reliable weighted estimates.

Table 22. 2010 CPEX Address Frame Accuracy and Quality: Weighted Census Unedited File New Adds by Original Source

Original Source	Unweighted CUF New Adds	Weighted CUF New Adds	Weighted Percent (SE[1])
**Total CUF New Adds ……..... **	**2,273**	**1,288,754**	**100.0**
Correct Enumerations (Correctly Added by Census) ……	1,623	960,527	74.5 (2.8)
Fall 2009 Delivery Sequence File/Spring 2010 Delivery Sequence File ……...	171	*	*
Group Quarters Validation……..…..	81	*	*
Update Leave/Urban Update Enumerate /Urban Update Leave…….....	406	285,459	22.2 (3.6)
Remote Alaska/Remote Update Enumerate/Update Enumerate……..	315	45,143	3.5 (1.0)
New Construction…….	91	*	*
Enumeration of Transitory Locations……..	161	*	*
Nonresponse Followup/Vacant Delete Check…….	356	235,277	18.3 (4.7)
Non-ID Operations……	30	*	*
Other…….	12	*	*
Erroneous Enumerations (Added-in-Error by Census)	650	328,227	25.5 (2.8)
Fall 2009 Delivery Sequence File/Spring 2010 Delivery Sequence File……	50	*	*
Group Quarters Validation…….....	136	*	*
Update Leave/Urban Update Enumerate /Urban Update Leave…….....	160	*	*
Remote Alaska/Remote Update Enumerate/Update Enumerate……..….	108	*	*
New Construction……..….	27	*	*
Enumeration of Transitory Locations……..…..	30	*	*
Nonresponse Followup/Vacant Delete Check……....	125	*	*
Non-ID Operations…….....	14	*	*
Other……	0	*	*

[1] Standard Error.
Source: CUF Adds Analysis Files.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table 23 shows the distribution of CUF Reinstated Adds by the Original Source variable. About 44.1 percent of the CUF Reinstated Adds correctly added by the census had the DSF as the original source and 34.1 percent had a pre-2010 Census operation as the original source.

Table 23. 2010 CPEX Address Frame Accuracy and Quality: Weighted Census Unedited File Reinstated Adds by Original Source

Original Source	Unweighted Count CUF Reinstated Adds	Weighted Count CUF Reinstated Adds	Weighted Percent (SE[1])
**Total CUF Reinstated Adds……..…… **	**2,496**	**2,312,356**	**100.0**
Correct Enumerations (Correctly Added by Census)……..	1,826	1,906,543	82.5 (3.3)
Pre-2010 Census Operations……	763	787,683	34.1 (8.6)
Pre-2010 Census Delivery Sequence Files……..…	928	1,018,630	44.1 (8.6)
Address Canvassing……	21	*	*
LUCA Appeals…….……....	114	*	*
Erroneous Enumerations (Added in Error by Census)…….	670	405,812	17.5 (3.3)
Pre-2010 Census Operations…….……..	258	161,416	7.0 (1.4)
Pre-2010 Census Delivery Sequence Files….....	248	*	*
Address Canvassing……	0	*	*
LUCA Appeals…….…...	164	*	*

[1] Standard Error.
Source: CUF Adds Analysis Files.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table 24 shows the added-in-error rate for the sample block clusters in FHU matching. Note that estimates in the table are unweighted because there is no appropriate weight available for the clusters in FHU themselves.

- A total of 1,153 block clusters had at least one CUF add record, which is 18.8 percent of the 6,148 FHU sample block clusters.
- Of the clusters having a CUF Add, about 55.9 percent had no records that the census added-in-error.
- About 17.9 percent of the clusters had an added-in-error rate of greater than zero and less than or equal to 50 percent.
- The remaining 26.2 percent of the clusters had an added-in-error rate of greater than 50 and less than or equal to 100 percent.
- Over one-fifth of clusters with an add action had an added-in-error rate over 91 percent.

These results suggest that the CUF Adds added-in-error were clustered geographically and were not distributed evenly throughout the FHU sample block clusters.

Table 24. 2010 CPEX Address Frame Accuracy and Quality: Block Cluster Level Added-in-Error Rate for CUF Adds from Final Housing Unit Matching

Error Rate within Cluster	Unweighted Number of Clusters	Unweighted Percent of Added-in-Error Records in Cluster*
Total clusters in sample with an add ……..……	**1,153**	**100.3[1]**
0 Percent - no adds in error……	644	55.9
1 to 50 Percent of adds in error…….…….	206	17.9
10 Percent or less of adds in error……...……..	17	1.5
11 to 20 Percent of adds in error……...……	23	2.0
21 to 30 Percent of adds in error……...……	23	2.0
31 to 40 Percent of adds in error……...……	49	4.3
41 to 50 Percent of adds in error……...……	94	8.2
51 to 100 Percent of adds in error……	303	26.2
51 to 60 Percent of adds in error……..…….	9	0.8
61 to 70 Percent of adds in error……..…….	18	1.6
71 to 80 Percent of adds in error……..…….	15	1.3
81 to 90 Percent of adds in error……...……	11	1.0
91 to 100 Percent of adds in error……...……..	250	21.7

[1] Percent does not round to 100 due to rounding error.
*Percent may not sum to applicable subtotals.
Source: CUF Adds Analysis Files.

5.5 Census Unedited File Deletes/Duplicates

This section answers the following research question:

By source/operation, what proportion of post-AC deleted/duplicated units were correctly deleted and erroneously deleted, according to their final CCM status?

A total of 19,152 CUF Deletes/Duplicates matched to the CCM sample of IHU matching results. This total includes records that CCM classified as GQs in IHU. Table 25 shows the 2010 Census unweighted and weighted estimates of CUF Deletes/Duplicates that the census correctly deleted and deleted-in-error according to the CCM combined IHU and FHU matching results. If a record was in FHU, the estimate used the FHU match code. However, if a record was only in IHU, the estimate used the IHU match code.

For comparison, the table includes the weighted estimates from the Census 2000 evaluation. As mentioned in the limitations, comparisons to Census 2000, particularly on weighted counts are difficult to make because of differences in both coverage measurement and decennial census procedures and methodology. Other possible explanations for the difference in the weighted counts include:

- The 2010 Census AC operation, with improvements to the procedures including a within operation delete validation, by incorporating automation, and by covering more of the country, may have resulted in a "cleaner" address inventory for the enumeration operations. In addition, the GQV was a new operation for the 2010 Census and likely improved the address inventory for the census enumeration operations. This improved inventory appeared to have resulted in less deletes and duplicates for the 2010 Census. The ratio of deletes/duplicates to the total number of housing units in the 2010 Census was 3.7 percent. As mentioned in the Background section, the ratio of deletes to housing units for Census 2000 was 7.4 percent.

- This evaluation used the CCM cluster weight (Weight2), and the Census 2000 evaluation used the census IHU weight (WeightC). Using WeightC in this analysis resulted in an overestimate of the number of CUF Deletes/Duplicates[15]. The cluster weight provided an estimate closer to the true count of CUF Deletes/Duplicates in this evaluation. However, even though the weights used in the two evaluations were different, the proportions should be comparable.

Despite the differences in the weighted counts, the weighted percentages between Census 2000 and the 2010 Census were similar. For the 2010 Census, the operations correctly deleted about 81.5 percent of the CUF Deletes/Duplicates, while Census 2000 correctly deleted 85.6 percent of the DMAF deletes. According to the CCM results, the 2010 Census deleted-in-error 18.5 percent of the CUF Deletes/Duplicates. In Census 2000, 14.4 percent of the DMAF deletes were incorrect. Since standard errors were not available from the Census 2000 evaluation, the DSSD cannot determine whether the difference in the percentages was statistically significant. However, the records the 2010 Census addresses deleted-in-error represented a weighted count of 900,369 addresses that the census may have omitted. This estimate of records deleted-in-error was about 26.6 percent less than the weighted estimate of 1.2 million housing units found by Smith et al. (2003).

[15] Using WeightC for this evaluation resulted in a weighted count of about 16.3 million, which is higher than the approximately 4.9 million national count of CUF Deletes/Duplicates.

Table 25. 2010 CPEX Address Frame Accuracy and Quality: Status of Census Unedited File Deletes/Duplicates – Combined Initial Housing Unit and Final Housing Unit Matching Results

CUF Delete/Duplicate Status	Census 2000		2010 Census		
	Weighted Count	Weighted Percent (SE[1])	Unweighted Count	Weighted Count	Weighted Percent (SE[1])
Total CUF Deletes/Duplicates …	**8,536,752**	**100.0**	**19,152**	**4,877,483**	**100.0**
Correctly Deleted …….....	7,309,409	85.6 (n/a)	16,694	3,977,115	81.5 (0.9)
Deleted-in-Error …….....	1,227,343	14.4 (n/a)	2,458	900,369	18.5 (0.9)

n/a is not available.
[1] Standard Error.
Source: Smith, et al, 2003 and CUF Delete/Duplicate Analysis File.

For the 2010 Census estimate in Table 25 above, the analysis used the FHU match code results for addresses in FHU. For addresses not in FHU, the estimates used the IHU match result. Although these results approximated the methodology used in the Census 2000 evaluation, the values for both Census 2000 and the 2010 Census likely overestimated the number and percentage of correctly deleted CUF delete/duplicate records. Three key factors influenced these estimates:

1. The results for both censuses assumed that deletes/duplicates matching to the CCM unresolved units were correctly deleted. A total of 246 cases were unresolved in IHU for the 2010 Census (see Table 27), but most of these were either not in FHU or became resolved. Some of these resolved cases had FHU match codes that make them "deleted-in-error" records.

2. Table 25 above assumed that the records classified as GQs in IHU for the 2010 Census were correct deletes. The IHU matching classified 225 records as GQs (see Table 27). Only 12 records remained classified as GQs after FHU matching. The rest were either not in FHU or changed their enumeration status. Some of the changed cases had FHU match codes that make them "deleted-in-error" records. The methodology for listing GQ in 2000 was significantly different from 2010. The Census 2000 evaluation completely excluded records classified by A.C.E. as GQs. As a result, the Census 2000 evaluation did not account for records that changed their status from a GQ in IHU to either erroneous or correct enumerations in FHU.

3. The Census 2000 estimates assumed that all the units classified in IHU as erroneous enumerations were correctly deleted by the census. However, some of these records were either not in FHU or changed their enumeration status in FHU to be correct enumerations and thus, were deleted-in-error by the census. Table 28 below examines these records in more detail and shows how many records changed their enumeration status from IHU to FHU.

To adjust for these factors, Table 26 below shows the status of the CUF delete/duplicate records according to the FHU matching results. (Note that estimates based on just the FHU matching use the WeightE variable.) These results indicate that the weighted percent of CUF Deletes/Duplicates correctly deleted was 74.2 percent and the weighted percent of CUF Deletes/Duplicates deleted-in-error was 25.8 percent.

Table 26. 2010 CPEX Address Frame Accuracy and Quality: Status of Census Unedited File Deletes/Duplicates – Final Housing Unit Matching Results

CUF Delete/Duplicate Status	Census 2000		2010 Census		
	Weighted Count	Weighted Percent (SE[1])	Unweighted Count	Weighted Count	Weighted Percent (SE[1])
Total CUF Deletes/Duplicates	**8,536,752**	**100.0**	**8,013**	**4,850,528**	**100.0**
Correctly Deleted……..	7,309,409	85.6 (n/a)	5,854	3,599,162	74.2 (1.8)
Deleted-in-Error……	1,227,343	14.4 (n/a)	2,159	1,251,366	25.8 (1.8)

n/a is not available.
[1] Standard Error.
Source: Smith, et. al, 2003 and CUF Delete/Duplicate Analysis File.

The tables in the rest of this section provide details on the "correctly deleted" and "deleted-in-error" estimates.

Table 27 shows the matching results from the IHU for the CUF Deletes/Duplicates in the CCM sample. Highlights from Table 27 include the following:

- Over 21 percent of the records did not have an IHU match code. These were records added by a census operation or were some other late add, such as LUCA appeals records, then subsequently, deleted from the census. IHU did not process records added after the January 2010 UCM.

- According to IHU, just over 41 percent of the CUF Deletes/Duplicates were "not housing units" on Census Day. This implies that about 41 percent of the time the census and CCM agreed, the census address was nonresidential or did not exist as a housing unit on Census Day. Since CCM classified these records as erroneous enumerations, the census correctly deleted the units.

- The IHU classified about 6.7 percent of the CUF Deletes/Duplicates as Census Duplicates. This means that the census address inventory had two or more records representing the same housing unit. Census Duplicates are erroneous enumerations and were counted as correctly deleted by the census.

- About 4.1 percent of the CUF Deletes/Duplicates were classified as census nonmatches in IHU. This means that CCM did not list and match a unit from the IL to the census address, but the Field Followup operation confirmed that the unit existed as a housing unit on Census Day. Some of these nonmatches are census geocoding error. In other words, the census geocoded the unit to the CCM block cluster, but the followup operation found the census address in a surrounding block. Table 30 examines the census nonmatches in more detail.

- The IHU matched almost 25 percent of the CUF Deletes/Duplicates. This means that the IHU found a CCM address matching the census address of the CUF Delete/Duplicate that existed as a housing unit on Census Day. Table 30 examines these matches in more detail.

Table 27. 2010 CPEX Address Frame Accuracy and Quality: Census Unedited File Deletes/Duplicates by CCM Initial Housing Unit Matching Results

CCM Enumeration Status	Unweighted CUF Deletes	Weighted CUF Deletes	Weighted Percent (SE[1])
Total CUF Deletes/Duplicates ……	**19,152**	**4,877,483**	**100.0**
Missing IHU Match Code……	3,757	1,051,860	21.6 (1.7)
Group Quarters…….	225	*	*
Erroneous Enumeration……..	9,626	2,386,831	48.9 (1.7)
Not a Housing Unit…….	8,005	1,998,653	41.0 (1.6)
Geocoding Error…….	21	*	*
Census Duplicates……..	1,354	324,547	6.7 (0.8)
Unresolved Units……	246	*	*
Correct Enumeration……	5,544	1,413,206	29.0 (1.4)
Census Nonmatches[2]……..…	731	199,503	4.1 (0.3)
Matches……..	4,813	1,213,703	24.9 (1.3)

[1] Standard Error.
[2] Includes census addresses geocoded to the sample block cluster but actually located in a surrounding block.
Source: CUF Delete/Duplicate Analysis File.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table 28 examines the FHU enumeration status of the CUF Deletes/Duplicates that IHU classified as erroneous enumerations. Some of these records changed their status from an erroneous enumeration (and correctly deleted by the census) to a correct enumeration in FHU (and deleted-in-error by the census). Details are as follows:

- A total of 35.6 percent of the records were not in FHU.

- An unweighted count of 90 CUF Deletes/Duplicates match to addresses that were linked to one or more duplicate records (i.e., CCM primaries). The analysis counted these records as correctly deleted.

- Of the records remaining in FHU, most of the records (55.8 percent) remained erroneous enumerations in FHU and were correctly deleted by the census.

- About 7.6 percent of the records became correct enumerations in the FHU processing and were deleted-in-error by the census.

Table 28. 2010 CPEX Address Frame Accuracy and Quality: FHU Status of Census Unedited File Deletes/Duplicates Classified as Erroneous Enumerations in Initial Housing Unit Matching

CCM Enumeration Status	Unweighted CUF Deletes	Weighted CUF Deletes	Weighted Percent (SE[1])
Total IHU Erroneous Enumerations……	**9,626**	**2,386,831**	**100.0**
Not in FHU……..……..……...	5,630	850,014	35.6 (2.3)
CCM Duplicates……	90	*	*
Not a Housing Unit…….…….	3,003	1,131,793	47.4 (2.1)
Geocoding Error…….…….	19	*	*
Census Duplicates……..……..	458	192,508	8.1 (1.3)
Unresolved…….……..	36	*	*
Correct Enumerations – Deleted-in-Error by Census….	390	180,358	7.6 (0.8)
Census Nonmatches……...	16	*	*
Census Matches……	374	176,519	7.4 (0.8)

[1] Standard Error.
Source: CUF Delete/Duplicate Analysis File.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table 29 examines the FHU enumeration status of the CUF Deletes/Duplicates coded as census nonmatches in IHU.

- About 60.3 percent of the CUF Deletes/Duplicates coded as census nonmatches in IHU had FHU match codes that contributed to the "deleted-in-error" estimate. The CCM field followup confirmed that these addresses existed as housing units on Census Day.

- About 29.2 percent of the IHU CUF Deletes/Duplicates that were census nonmatches were not in FHU.

- An unweighted count of 16 of the IHU census nonmatches were primaries to census duplicates. This means that the CUF delete/duplicate matched to an address that is linked to one or more duplicate census units. As mentioned in the methodology section, the analysis counted these cases as correctly deleted.

- An unweighted count of 48 of the IHU census nonmatches had a FHU match code of census geocoding error. The census geocoded these addresses to the CCM sample block cluster, but CCM found the addresses in a surrounding block. The analysis assumed that the census correctly deleted these addresses.

Table 29. 2010 CPEX Address Frame Accuracy and Quality: FHU Status of Census Unedited File Deletes/Duplicates that were Census Nonmatches in IHU

CCM Enumeration Status	Unweighted CUF Deletes	Weighted CUF Deletes	Weighted Percent (SE[1])
Total Census Nonmatches…….	**731**	**199,503**	**100.0**
Not in FHU…….	359	58,319	29.2 (3.5)
CCM Primaries……	16	*	*
Unresolved……...	1	*	*
Duplicates…….....	4	*	*
Not a Housing Unit…….	5	*	*
Group Quarters……	1	*	*
Census Nonmatches – Geocoding Error…….	48	*	*
Deleted- in-Error……...	297	120,211	60.3 (3.5)

[1]Standard Error.
Source: CUF Delete/Duplicate Analysis File.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table 30 examines the FHU enumeration status of CUF delete/duplicate addresses coded as census matches in IHU.

- About 45.3 percent of the IHU census matches had a FHU match code counted in this analysis as deleted-in-error.
- Of the correctly deleted addresses, the FHU determined that 11.3 percent did not exist as a housing unit on Census Day.
- About 34.0 percent of the census matches were not in FHU.

Table 30. 2010 CPEX Address Frame Accuracy and Quality: FHU Status of Census Unedited File Deletes/Duplicates that were Census Matches in IHU

CCM Enumeration Status	Unweighted CUF Deletes	Weighted CUF Deletes	Weighted Percent (SE[1])
Total Census Matches……...	**4,813**	**1,213,703**	**100.0**
Not in FHU……..…….	2,725	412,888	34.0 (2.2)
CCM Primaries…….	197	*	*
Duplicates…….	40	*	*
Unresolved……	8	*	*
Geocoding Error……...	5	*	*
Not a Housing Unit……..	366	137,567	11.3 (1.3)
Census Nonmatch…….	5	*	*
Deleted- in-Error……	1,467	549,581	45.3 (1.9)

[1] Standard Error.
Source: CUF Delete/Duplicate Analysis File.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table 31 shows results from the NPC staff's investigation into a subset of the CUF Deletes/Duplicates that IHU coded as deleted-in-error. While these results are anecdotal, the table is instructive in indicating reasons why the census may have deleted the addresses in error. The top three reasons why census deleted addresses in error are as follows:

- The census classified about 21.5 percent of the addresses as uninhabitable, but CCM enumerators disagreed and classified the units as habitable.

- About 19.4 percent of the addresses deleted-in-error were units within a multi-unit structure.

- For about 16.0 percent of the addresses, the NPC staff determined that the census made an error for an unknown reason.

Table 31. 2010 CPEX Address Frame Accuracy and Quality: IHU Census Unedited File Deletes/Duplicates Deleted-in-error by Reason for Delete Error

Reason for Delete Error	Unweighted CUF Deletes	Unweighted Percent
Total……	**1,724**	**100.0**
Issues with Housing Unit Definition		
Unit is new construction……	11	0.6
Unit that Census thought was not habitable (boarded up, etc.), but CCM has as habitable……..…	371	21.5
Unit is seasonal housing……..……	79	4.6
Issues with Overlooking Addresses or Locating Addresses		
Building is a multi-unit that was completely missed even if other buildings in the complex were found……..……	80	4.6
Record has a map and/or geography error…….…..……	7	0.4
Unit is a hidden unit that is a separate structure…….……	171	9.9
The same MAFID represents an address with significant changes…….	27	1.6
Unit has slightly different address, such as house number, street name, etc……..	249	14.4
Unit is in a trailer park and it is distinct from a hidden unit situation……..	105	6.1
Unit is in a large or small multi-unit where the building was not missed but the unit or units were……...	335	19.4
The record was deleted from the wrong block and never moved or added to the correct block……..	2	0.1
Other		
Unit deleted due to lister error or an unknown reason……..……	275	16.0
Unit was a duplicate but no primary can be found…….….	12	0.7

Source: CUF Delete/Duplicate Analysis File.

Table 32 shows the distribution of CUF Deletes/Duplicates by census region according to FHU matching results.

- According to the FHU matching results, the South region had the highest percentage of correctly deleted units (31.1 percent) and the highest percentage of units deleted-in-error (11.9 percent).

Table 32. 2010 CPEX Address Frame Accuracy and Quality: Status of Census Unedited File Deletes/Duplicates in CCM Final Housing Unit Matching Results by Census Region

Status of CUF Deletes/Duplicates by Region	Unweighted CUF Adds	Weighted CUF Adds	Weighted Percent (SE[1])
Total CUF Deletes/Duplicates……..	**8,013**	**4,850,528**	**100.0**
Correctly Deleted…….....	5,854	3,599,162	74.2 (1.8)
Midwest……..	938	675,338	13.9 (2.4)
Northeast…….	992	773,022	15.9 (2.3)
South…….	2,111	1,508,103	31.1 (2.5)
West…….	1,813	642,699	13.3 (1.6)
Deleted-in-Error…….....	2,159	1,251,366	25.8 (1.8)
Midwest……	387	268,823	5.5 (0.7)
Northeast……..	356	234,362	4.8 (0.5)
South……..	713	578,686	11.9 (1.4)
West……..	703	169,495	3.5 (0.5)

[1] Standard Error.
Source: CUF Delete/Duplicate Analysis Files.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table 33 shows the deleted-in-error rate for the sample block clusters in FHU matching. Note that estimates in the table are unweighted because there is no appropriate weight available for the FHU sample block clusters.

- A total of 2,980 block clusters had at least one CUF delete/duplicate record, which is about 48.5 percent of the 6,148 block clusters in FHU.
- Of the clusters having a CUF Delete/Duplicate, about 61.6 percent had no records that the census deleted-in-error.
- About 23.5 percent of the clusters had a deleted-in-error rate of greater than zero and less than or equal to 50 percent.

- The remaining 14.9 percent of the clusters had a deleted-in-error rate of greater than 50 and less than or equal to 100 percent.

These results suggest that the CUF Deletes/Duplicates deleted-in-error were clustered geographically and were not distributed evenly throughout the FHU sample block clusters.

Table 33. 2010 CPEX Address Frame Accuracy and Quality: Block Cluster Level Deleted-in-error Rate for CUF Deletes/Duplicates from Final Housing Unit Matching

Error Rate within Cluster	Unweighted Number of Clusters	Unweighted Percent of Deleted-in-Error Records in Cluster
Total clusters in sample with a delete/duplicate …….	**2,980**	**100.0**[1]
0 Percent - no delete/duplicates in error……..……	1,835	61.6
1 to 50 Percent of delete/duplicates in error……...…….	700	23.5
10 Percent or less of delete/duplicates in error……..……..	94	3.2
11 to 20 Percent of delete/duplicates in error……..……	182	6.1
21 to 30 Percent of delete/duplicates in error……...……	95	3.2
31 to 40 Percent of delete/duplicates in error……...…….	138	4.6
41 to 50 Percent of delete/duplicates in error……..…….	191	6.4
51 to 100 Percent of delete/duplicates in error……..…….	445	14.9
51 to 60 Percent of delete/duplicates in error…….	20	0.7
61 to 70 Percent of delete/duplicates in error……..…….	71	2.4
71 to 80 Percent of delete/duplicates in error……..……	19	0.6
81 to 90 Percent of delete/duplicates in error……...……	8	0.3
91 to 100 Percent of delete/duplicates in error……..…...	327	11.0

[1] Percent does not add to 100 due to rounding error.
Source: CUF Deletes/Duplicates Analysis File.

The 2010 Census did not have a specific Kill Process operation or Housing Unit Unduplication Operation like Census 2000. Yet, the 2010 Census retained the standard from Census 2000 that before classifying a record as deleted from the census, it must have double validation that the record is a delete. That is, two operations or phases with in operation must confirm the delete status of the record. Thus, it is difficult to determine from the available data on the files, the specific operation or source of a CUF delete/duplicate record. However, since Nonresponse Followup (NRFU) was a major contributor to the census deletes, Table 34 and Table 35 show the distribution of CUF Deletes/Duplicates by the NRFU Data Capture Status variable (NRSTATUS2) and by the NRVS variable, respectively.

In Table 34, of the CUF Deletes/Duplicates, two NRFU Data Capture Status categories stand out as being deleted-in-error at a higher rate. With 7.4 percent of the CUF Deletes/Duplicates, the uninhabitable records represented 28.8 percent of the records deleted-in-error. The records classified as Demolished were 6.1 percent of all the CUF Deletes/Duplicates and 23.5 percent of the records deleted-in-error.

Table 34. 2010 CPEX Address Frame Accuracy and Quality: Status of Census Unedited File Deletes/Duplicates in Final Housing Unit Matching by Nonresponse Followup Data Capture Status

Enumeration Status/Non-Response Followup Status	Unweighted CUF Deletes	Weighted CUF Deletes	Weighted Percent (SE[1])
Total CUF Deletes/Duplicates……	**8,013**	**4,850,528**	**100.0**
Correctly Deleted…….	5,854	3,599,162	74.2
Not Captured in Nonresponse Followup…….	677	433,208	8.9 (1.0)
Duplicate……..	700	445,449	9.2 (1.5)
Demolished……..	2,125	1,353,514	27.9 (2.2)
Empty Mobile Home/Trailer……..……	405	262,613	5.4 (1.0)
Nonresidential……..	921	579,075	11.9 (2.3)
Occupied……..	101	*	*
Uninhabitable…….……..	818	427,165	8.8 (1.4)
Vacant – Regular…….…….	87	*	*
Vacant - Usual Home Elsewhere……	20	*	*
Deleted-in-Error…….	2,159	1,251,366	25.8 (1.8)
Not Captured in Nonresponse Followup……..……..	260	*	*
Duplicate……..	178	*	*
Demolished……..	474	293,833	6.1 (0.5)
Empty Mobile Home/Trailer……	265	*	*
Nonresidential……..	272	*	*
Occupied……..	27	*	*
Uninhabitable……...	615	360,533	7.4 (1.0)
Vacant – Regular…….…….	62	*	*
Vacant - Usual Home Elsewhere……	6	*	*

[1] Standard Error.
Source: CUF Delete/Duplicate Analysis File.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table 35 shows the distribution of CUF Deletes/Duplicates by the NFRUV Status variable. Of the records deleted-in-error, one category stands out with the highest error rate. Records classified as Undeliverable as Addressed (UAA) that went to NRFU and were deletes in NRFU were 15.6 percent of the CUF Deletes/Duplicates and 60.4 percent of the records deleted-in-error.

Table 35. 2010 CPEX Address Frame Accuracy and Quality: Status of CUF Deletes/Duplicates in Final Housing Unit Matching by Non-Response Followup Vacancy Status

Enumeration Status/Non-Response Followup Status	Unweighted CUF Deletes	Weighted CUF Deletes	Weighted Percent (SE[1])
**Total CUF Deletes/Duplicates……..…….. **	**8,013**	**4,850,528**	**100.0**
Correctly Deleted……..……..	5,854	3,599,162	74.2 (1.8)
Not in NRFU……...……..……	2,744	1,502,589	31.0 (2.9)
Delete in NRFU - verified as a delete by VDC[2]……..…	601	404,000	8.3 (0.8)
Delete in NRFU - vacant in VDC…….	1	*	*
Delete in Update/Leave - vacant in NRFU (not VDC eligible)……..…..	3	*	*
Delete in Update/Leave - Delete in NRFU (not VDC eligible)……..……..	379	263,373	5.4 (0.8)
UAA[3] that went to NRFU - Delete in NRFU (not VDC eligible)……..……..	2,121	1,425,738	29.4 (2.5)
UAA that went to NRFU - vacant in NRFU…….	5	*	*
Deleted-in-Error……..…….	2,159	1,251,366	25.8 (1.8)
Not NRFU or VDC ……..…	740	236,511	4.9 (0.5)
Delete in NRFU - verified as a delete by VDC……..…	249	*	*
Delete in NRFU - vacant in VDC…….	0	*	*
Delete in Update/Leave - vacant in NRFU (not VDC eligible)…….....	0	*	*
Delete in Update/Leave - Delete in NRFU (not VDC eligible)…….....	119	*	*
UAA that went to NRFU - Delete in NRFU (not VDC eligible)……..……..	1,050	756,220	15.6 (1.4)
UAA that went to NRFU - vacant in NRFU…….	1	*	*

[1] Standard Error.
[2] Vacant Delete Check.
[3] Undeliverable as Addressed.
Source: CUF Delete/Duplicate Analysis File.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

5.6 Geocoding Error

Geocoding results

This section answers the following research question: **What was the total estimated percentage of census addresses geocoded to the incorrect 2010 Census collection block?**

This evaluation attempted to search for records missing from census in an extended search area, called a "buffer," up to five km from the boundary of the block cluster. CCM only matched their IL to census addresses in the first ring of blocks around the block cluster. This buffer search area

traced around the shape of each cluster. For more information on the buffer search area refer to Johnson (2011a).

As Table 36 shows, it appears the geocoding error rate is decreasing as time progresses. In the 1998 MAF Quality Improvement Program, the geocoding error rate was reported as 6.2 percent; in Census 2000, it was found to be 4.8 percent (Ruhnke, 2003); and in the 2010 Census, it was 1.5 percent. Without a standard error available from 1998, a confidence interval cannot be calculated. However, the Census 2000 and 2010 Census geocoding error rates had mutually exclusive confidence interval, which suggests a statistically significant difference. The CCM geocoding error rate, at 1.3 percent, actually measures a different variation of geocoding error than this evaluation. *Errors of inclusion* are where an address was listed as being in the CCM sample and actually existed beyond the surrounding ring of blocks. This evaluation measured only *errors of exclusion* where the address was listed beyond the CCM sample block clusters when it actually existed in the sample blocks.

Table 36. 2010 CPEX Address Frame Accuracy and Quality: Status of Geocoding in 1998, 2000, and 2010

Geocoding Error Rates	Weighted Percent	95 Percent Confidence Interval Lower	95 Percent Confidence Interval Upper	SE[1]
1998 MAF QIP Geocoded in Error Rate…….	6.2	**n/a**	**n/a**	**n/a**
2000 Geocoded in Error Rate…….	4.8	4.2	5.4	0.3
2010 CCM Geocoded in Error Rate *(Errors of Inclusion)*……	1.3	1.1	1.5	0.1
2010 AFAQ Geocoded in Error Rate*(Errors of exclusion)*……..	1.5	1.3	1.7	0.1

n/a not available.
[1] Standard Error.
Source: Ruhnke 2003, Keller and Fox 2012, Geocoding Analysis File .

Did the geocoding error estimate vary by type of enumeration area or by census region?

Table 37 presents the geocoding error estimate within each census region. Presenting the within region error rate, as opposed to the percent of all records, by geocoding status, controls for the large size of the South region.

- Even after adjusting for the large number of records in the South, the within region geocoding error rate was the highest, at 2.3 percent.

- The lowest geocoding error rate was in the Midwest at 0.7 percent.

Table 37. 2010 CPEX Address Frame Accuracy and Quality: Geocoding Error Rate within Census Region

Status of Geocoding by Region	Weighted Erroneously Geocoded Count	Weighted Correctly Geocoded Count	Weighted Percent Within Region (SE[1])
Total Records……	**1,992,600[2]**	**130,543,009**	**1.5 (0.1)**
Midwest……..	206,175	28,743,504	0.7 (0.2)
Northeast……..…..	301,871	23,957,852	1.2 (0.2)
South……..……..…….	1,191,849	50,447,692	2.3 (0.3)
West…….……..	292,706	27,393,961	1.1 (0.2)

Source: Geocoding Analysis File.
[1] Standard Error. [2] Subtotals may not sum due to rounding error.

Table 38 presents the within TEA geocoding error rate for Mailout/ Mailback and Update/Leave areas. There were not enough cases to produce reliable estimates for the Update/Enumerate and Unknown TEAs. The geocoding error rate within Mailout/ Mailback was the same as the national rate, at 1.5 percent, and geocoding error was slightly higher within Update/Leave areas at 1.9 percent.

Table 38. 2010 CPEX Address Frame Accuracy and Quality: Geocoding Error Rate by Type of Enumeration Area

Status of Geocoding by TEA	Weighted Correctly Geocoded Count	Weighted Erroneously Geocoded Count	Weighted Percent Within TEA (SE[1])
Total Records……	**130,543,009[2]**	**1,992,600**	**1.5 (0.1)**
Mailout/ Mailback……..……..	118,383,050	1,752,323	1.5 (0.1)
Update/Leave…….……	10,780,391	209,750	1.9 (0.4)
Update/Enumerate…….……	1,274,600	*	*
Unknown TEA…….…….	105,985	*	*

[1] Standard Error. [2] Subtotals may not sum due to rounding error.
Source: Geocoding Analysis File. *There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table 39 examines where "missing" census records were mistakenly geocoded, relative to their true location on the ground.

- This evaluation included geocoding error within the cluster and up to five km from the cluster, which found a geocoding error rate of 1.5 percent.

- Beyond one km, only 28 unweighted records were found.

Table 39. 2010 CPEX Address Frame Accuracy and Quality: Search Area Location for Geocoding Error

Block Location Relative to CCM Cluster	Unweighted Count	Weighted Count	Weighted Percent (SE[1])
Total Misgeocoded Records……..……	**6,714**	**1,992,600**	**100.0**
Within Cluster & ring of surrounding blocks……	5,731	1,828,533	1.38 (0.1)
Between ring of surrounding blocks & 1 KM……...	955	156,322	0.12 (0.03)
Between 1 KM & 3 KM……...	22	*	*
Between 3 KM & 5KM……	6	*	*

[1] Standard Error.
Source: Geocoding Analysis File.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table 40 shows a relatively large disparity in the geocoding error rate by original source. The lowest geocoding error rate was from records with a source from a census operation prior to the 2010 Census AC at one percent. On the other extreme, records added by 2010 Census AC, LUCA, or 2010 Census operations have a one percent higher geocoding error rate, at 2.5 percent, than the national average.

Table 40. 2010 CPEX Address Frame Accuracy and Quality: Geocoding Error Rate by Original Source

Status of Geocoding by Original Source	Weighted Correctly Geocoded Count	Weighted Erroneously Geocoded Count	Weighted Geocoding Error Rate By Source (SE[1])
Total Records……	**130,544,026[2]**	**1,992,600**	**1.5 (0.1)**
Pre- 2010 Census ……..…….	118,383,050	1,062,926	1.0 (0.2)
DSF…….……..	10,780,391	209,750	1.9 (0.3)
2010 AC/ LUCA/Census Operations..……..	5,973,072	156,225	2.5 (0.4)
Unknown Source……...	123,575	*	*

[1] Standard Error. [2] Subtotals may not sum due to rounding error.
Source: Geocoding Analysis File.*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table 41 presents the geocoding error rate by ACT. The ACT was assigned at a collection block level prior to AC, so blocks thought to be empty are not assigned an ACT for this evaluation. For this report, Unknown blocks, Non-Residential blocks, and Non-City-Style blocks (blocks that contain less than 70 percent city style addresses, i.e., PO Boxes, location descriptions) are combined into one category. The remaining categories are City-Style and Mixed. (Mixed includes blocks that are at least 70 percent city-style).

- Over 100 million records were in blocks with a City-Style ACT. The geocoding error rate of these records was similar to the national average at 1.4 percent.
- Mixed blocks had a slightly higher geocoding error rate at 1.9 percent.
- The worst geocoding error rate was for ACTs that are Unknown, Non-Residential, or Non-City-Style at 2.8 percent.

Table 41. 2010 CPEX Address Frame Accuracy and Quality: Geocoding Error Rate by Address Characteristics Type

Status of Geocoding by ACT of Collection Block	Weighted Correctly Geocoded Count	Weighted Erroneously Geocoded Count	Weighted Geocoding Error Rate By ACT (SE[1])
Total Records……	**130,544,026**	**1,992,600**	**1.5 (0.1)**
City-Style Blocks…….	101,199,397	1,403,742	1.4 (0.2)
Mixed Blocks (City and Non-City-Style)….	26,352,255	504,288	1.9 (0.3)
Unknown/Nonresidential/Non-City-Style…	2,992,374	84,570	2.8 (0.6)

Source: Geocoding Analysis File.
[1] Standard Error.

This report was unable to answer the following research question: **How many addresses coded as "missing" from the Census by CCM were on the MAF but excluded from the census?**

In Census 2000, a weighted count of 1.3 million housing units were erroneously excluded from Census due to the DMAF filter. However, to prevent the filter from causing a similar loss of potentially valid units, the 2010 Census AC filter was designed to allow census to validate questionable units.

Records were excluded from AC for a limited number of reasons:

a) They did not contain a collection block and therefore, would be ineligible for the block-based operation;
b) They did not meet minimal address requirements, such as having a house number and street name or location description with a map spot;
c) They were previously identified as duplicates of records eligible for AC and the census;
d) They were sent to other operations such as sensitive group quarters or military vessels.

This evaluation was unable to include these excluded records in the matching universe for many of the same reasons why AC excluded them. For example, if they did not contain a collection block, the authors could not load them into the search area, or if a record has incomplete address information, it will be difficult to match.

6 Related Evaluations, Experiments, and Assessments

The following 2010 Census Evaluations, Experiments, and Assessments are related to this evaluation:

- 2010 Census Address Canvassing Operational Assessment (AL OIT, 2011)
- 2010 CPEX Evaluation of Address List Maintenance Using Supplemental Data Sources (Thomasewski, *Forthcoming 2013*)
- 2010 Census Evaluation to Assess Effect of Census Coverage Measurement Search Area and Census Address List Formation Rules on CCM Estimates (Bray, 2012)
- 2010 Census Components of Census Coverage for Housing Units in the United States (Keller and Fox, 2012)

7 Conclusions and Recommendations

7.1 Conclusions

As the first major field operation of the census, it was important for AC to provide an accurate address inventory for the census enumeration operations. An accurate inventory reduces census costs and lessens the risk of either omissions from the census or an over-count. The results of this evaluation indicate that the AC operation was highly successful in accurately deleting the delete/duplicate records with the percentage of deleted-in-error addresses at 4.3 percent. This represents a weighted count of 786,294 addresses deleted-in-error and about 0.5 percent of all records processed in the AC operation.

Although the AC added records had a higher error rate (16.4 percent), this was understandable, since the AC procedures encouraged Listers to add addresses when in doubt about their status or potential future status. In addition, one of the objectives of AC was to add any potential Other Living Quarters for the GQV operation to classify and process. Even so, the weighted count of about 1.7 million records added-in-error was just 1.2 percent of all records processed in AC.

Both AC and the GQV operations were successful in providing a "cleaner" address inventory for the census enumeration operations. Compared to Census 2000, the CUF for the 2010 Census had fewer added addresses and deletes/duplicate records. As a ratio of added addresses to valid housing units, the 2010 Census ratio of 2.5 percent was lower than the 3.3 percent from Census 2000. The difference in delete/duplicate records was even greater. The 2010 Census ratio of delete/duplicate records to valid housing units was about 3.7 percent, while the Census 2000 ratio was about 7.4 percent.

Despite the difference in total CUF added addresses and CUF deleted/duplicated records, the percentage of addresses added-in-error appeared to be roughly similar between the 2010 Census and Census 2000 while the percentage of addresses deleted-in-error appeared to be higher in the 2010 Census. For the 2010 Census, about 20.4 percent of the CUF Adds were added-in-error,

while for Census 2000, about 16.1 percent of the HCUF Adds were added-in-error. For the CUF Deletes/Duplicates, the 2010 Census had about 25.8 percent deleted-in-error, while Census 2000 had 14.4 percent. As mentioned in the results, standard errors were not available from the Census 2000 evaluation, so the DSSD could not determine whether these differences were statistically significant. However, it is possible to conclude that the 2010 Census enumeration operations did not improve the error rates found in Census 2000 and the percentage of addresses deleted-in-error was appeared to be higher.

The errors for both AC and CUF addresses added-in-error and deleted-in-error tended to be clustered within a subset of the CCM sample block clusters. For AC, about 50 percent of the block clusters containing at least one added address had no errors, and 80 percent of the block clusters containing at least one delete/duplicate record had no errors. For the CUF, about 55.9 percent of the block clusters containing at least one added address had no errors, and 61.6 percent of the block clusters containing at least one delete/duplicate record had no errors. These results imply that specific blocks either had problem addresses, or the block itself had characteristics that presented problems for accurately identifying addresses.

The geocoding error rate improved between Census 2000 and the 2010 Census. In Census 2000, the geocoding error rate was 4.8 percent, while the error rate for the 2010 Census was 1.5 percent. A low geocoding error rate is important for accurately enumerating people in the correct location. It also allows surveys and future census operations to reduce costs in attempting to locate addresses. The lower geocoding error implies that the collection of GPS coordinates in AC and other geocoding activities conducted by the Census Bureau throughout the decade have been successful.

In summary:

- The deleted-in-error rate for AC delete/duplicate records of 4.3 percent was low. These records deleted-in-error represented about 0.5 percent of all records processed in the AC operation.

- The AC added addresses had an error rate of 16.4 percent. This represented about 1.2 percent of all records processed in the AC operation.

- The 2010 Census enumeration operations had fewer added addresses and deleted records compared to Census 2000, which suggests that AC and GQV operations provided a relatively more accurate address inventory to the enumeration operations.

- The CUF added-in-error rate of 20.4 percent for the 2010 Census were similar to the error rate found in Census 2000 while the CUF deleted-in-error rate of 25.8 percent for the 2010 Census was higher than the rate found in Census 2000.

- The AC and CUF addresses added-in-error and deleted-in-error were clustered in a subset of the CCM block clusters.

- The geocoding error rate significantly improved from 4.8 percent in Census 2000 to 1.5 percent in the 2010 Census.

7.2 Recommendations

- **Conduct research into the characteristics and predictors of added and deleted addresses.**

  Understanding the characteristics/predictors of change to the census address inventory is an important step in maintaining an accurate and high quality address frame. Some characteristics are well known, such as areas that experience new construction, but predictors of changes in the address inventory have received less study. Research using the 2010 Census Planning Database, Administrative Records, the Partnership Program, and the 2020 Census field tests can provide information to predict future changes. Areas with added and deleted addresses are likely to be spatially/geographically clustered. This finding is consistent with Boies, (2012) and has important implications for future decisions on targeting for listing operations.

- **Conduct research into the characteristics and predictors of addresses added-in-error and deleted-in-error.**

  Understanding where and why the census operations make errors in either adding or deleting addresses is important for an accurate census. The Census Bureau should conduct additional research into the addresses that the census added and deleted-in-error to determine characteristics of these addresses or of their blocks. Research using the 2010 Census Planning Database, Administrative Records, the Partnership Program and the 2020 Census field tests could provide information on specific addresses or blocks that need targeted training, procedures, or updating.

- **Weigh the costs and benefits of further decreasing geocoding error.**

  The results of this evaluation indicate that the geocoding error was significantly less than in Census 2000 and the authors recommend that the Census Bureau examine the costs and benefits associated with further decreasing geocoding error. Given a geocoding error rate of only 1.5 percent, it may not be worth implementing large-scale field operations only to improve geocoding. However, several low cost options may be available to continue to improve geocoding on the Master Address File. According to Tomaszewski (*Forthcoming 2013*), the majority of the Geographic Information Systems geocodes collected from local governments were accurate. Continued collection of these Geographic Information Systems data may lead to a smaller number of missing, ungeocoded, and misgeocoded units in the address frame, which translates into a higher quality address frame (reduced undercoverage and increased accuracy).

In addition to analyzing the costs and benefits associated with improving geocoding error, the authors also recommend that the current CCM search area is sufficient to measure geocoding error. Since CCM already matches records within one ring of surrounding blocks it may not be necessary to perform a separate evaluation, independent of the 2020 Census post enumeration survey measure, assuming producing this estimate is in line with the goals and design of the 2020 Census post enumeration survey program.

- **Fully integrate the housing unit added- and deleted-in-error statistics into the 2020 CCM housing unit studies.**

  In this evaluation, in order to calculate the housing unit added- and deleted-in-error statistics for the final 2010 Census, the 2010 CCM results were heavily leveraged. Additionally, very similar headquarters and field procedures were used to determine the 2010 AC added- and deleted-in-error statistics presented here. With these facts, combined with the knowledge that 2010 CCM staff have and continue to conduct very similar work, it may be more economical for the Agency to integrate these added- and deleted-in-error statistics, both post-listing and final census, into the 2020 Census post enumeration survey program. With increased funding and priority over the 2020 Census Evaluation program, the 2020 Census post enumeration survey program may be better suited to conduct the fieldwork and subsequent analyses in a timely manner, without the need to design, plan, and execute costly supplemental field operations.

8 Acknowledgements

This evaluation is the result of the coordinated efforts of numerous people in several areas of the Census Bureau. In DSSD the support and work of several managers including Kevin Shaw, Jennifer Reichert, and David Whitford, made this evaluation possible despite many resource and timing conflicts. In Geography Division, the authors wish to acknowledge and thank the staff who was instrumental in the creation of the Surrounding Block File. Several people on the Census Automation Staff in the TMO modified the CMOCS software for use in the Geocoding Error portion of the evaluation. Many people in the Field Division and the RO provided support for the two field operations conducted for this evaluation. Specifically, the authors wish to point out the efforts of staff in the Field Coverage Measurement Branch and the Field Labor and Crimes Branch. Somehow, they managed to fit this evaluation work into their busy schedules of CCM and current survey work. In addition, the CCM supervisors and CCM enumerators who participated in this evaluation worked hard to complete their workloads on a tight schedule. The CCM Section in the NPC also deserves special mention. Even though they were busy with CCM, they found time to answer questions, research selected cases, and conduct the clerical matching for this evaluation's two field operations. Staff in the Documents Services Branch of the NPC provided printing, form assembly, and shipping support for the field operations. Many people in the DSSD who worked on CCM provided support in various forms for this evaluation. The authors thank all the staff in these areas that made this evaluation possible.

9 References

Address List Operations Implementation Team (2011), "2010 Census Address Canvassing Operational Assessment," 2010 Census Planning Memorandum No. 168, January 10, 2012.

Boies, John L. (2012), "Final Report for the 2010 Census Evaluation of Address Canvassing Targeting and Cost Reduction," 2010 Census Program for Evaluations and Experiments, #A-09, July 12, 2012.

Bray, Rachel and Colt Viehdorfer (2012), "Census Coverage Measurement Estimation Report: Missing Data for Components of Census Coverage," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-G-09, May 22, 2012.

Bray, Rachel (2012), " Final Report for the 2010 Census Program for Evaluations and Experiments: Evaluation to Assess Effect of Census Coverage Measurement (CCM) Search Area and Census Address List Formation Rules on CCM Estimates," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-O-A-53, September 24, 2012.

Burchman, Joseph A. (2002), "Block Canvassing Operation," U.S. Census Bureau, April 5, 2002, page i.

Cantu, Aaron B. (2009), "Software Requirements Specification for the 2010 Census Coverage Measurement Obtain, Prepare, & Deliver Universe Control and Management File Software," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-D4-05R1, October, 14, 2009.

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.

Cronkite, Diane (2012), "Paper of the Census Coverage Measurement Initial Housing Unit Matching Activities," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-I-21, April 30, 2012.

Davis, Peter P. (2012), "2010 Census Coverage Measurement: Operational Summary of the Sample Design," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-C-27, February 21, 2012.

Dixon, Kelly, Melissa Blevins, Robert Colosi, Amanda Hakanson, Nancy Johnson, Karen Owens, Matt Stevens, Christine Gibson Tomaszewski (2008), "2008 Census Dress Rehearsal Address Canvassing Assessment Report," DSSD 2008 Census Dress Rehearsal Memorandum Series #D-20, April 15, 2008.

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.

Gbur, Philip M. (2010), "Transmittal of 2010 Census Technical Documentation '2010 Census: Operational Overview and Accuracy of the Data' Section", DSSD 2010 Decennial Census Memorandum Series #G-22, September 24, 2010.

Holland, Jonathan P. and Matthew Virgile, (2009), "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," 2010 Census Program for Evaluations and Experiments #A-05, July 24, 2012.

Imel, Jerry (2010), "Customer Requirements for the Address Filter for the 2010 Census Enumeration and Nonresponse Followup Address Products," DSSD 2010 Decennial Census Memorandum Series #D-09-R1, March 18, 2010.

Imel, Jerry and Diane Probst (2012), "Interface Control Document for the Data Exchanges between DSSD CEE and NPC MaRCS in Support of the HUE Operation Version 1.00," DSSD 2010 Decennial Census Memorandum Series, #H-35, June 30, 2011.

Contreras, Graciela, Diane Cronkite, Lora Rosenberger, Anne Wakim,(2012), "2010 Census Coverage Measurement Initial Housing Unit Independent Listing, Matching, and Followup Operations Assessment Report," 2010 Census Planning Memoranda Series No. 178-R, Sep 5, 2012.

Johnson, Nancy (2010), "Specifications for the 2010 Census Evaluation of Address Frame Accuracy and Quality Address Canvassing Delete and Duplicate Address Before Followup Clerical Matching," DSSD 2010 Decennial Census Memorandum Series #O-A-20, June 30, 2010.

Johnson, Nancy (2011a), "Customer Requirements for the 2010 Surrounding Collection Tract and 2010 Block Cluster Buffer Files," DSSD 2010 Decennial Census Memorandum Series #O-A-11, March 1, 2011.

Johnson, Nancy (2011b), "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 (2011c), "Processing Requirements for the 2010 Census Evaluation of Address Frame Accuracy and Quality – Geocoding Error Operation," DSSD 2010 Decennial Census Memorandum Series #O-A-34, June 22, 2011.

Johnson, Nancy (2012a), "2010 Census Evaluation of Address Frame Accuracy and Quality After Followup Clerical Matching Procedures for the Geocoding Error Operation," DSSD 2010 Decennial Census Memorandum Series #O-A-39, June 28, 2012.

Johnson, Nancy R., Alicia Green and Tanisha Willis, (2012b), " 2010 Census Housing Unit Evaluation Field Followup Procedures Training Materials for the 2010 CPEX Evaluation of Address Frame Accuracy and Quality," DSSD 2010 Decennial Census Memorandum Series #O-A-38, July 3, 2012.

Johnson, Nancy (2012c), "Requirements for the 2010 Evaluation of Address Frame Accuracy and Quality Housing Unit Evaluation Matching and Review Coding System," DSSD 2010 Decennial Census Memorandum Series #O-A-52, August 2, 2012.

Keller, Andrew and Tyler Fox (2012), “2010 Census Coverage Measurement Estimation Report: Components of Census Coverage for Housing Units in the United States,” DSSD 2010 Census Coverage Measurement Memorandum Series #2010-G-06, May 22, 2012.

Kephart, Kathleen M. (2012a), " Specifications for the 2010 Census Evaluation of Address Frame Accuracy and Quality Address Canvassing Delete and Duplicate Address After Followup Clerical Matching," DSSD 2010 Decennial Census Memorandum Series #O-A-49, November 9,2012.

Kephart, Kathleen M. (2012b), " 2010 Census Evaluation of Address Frame Accuracy and Quality Address Canvassing and Census “Deleted in Error” Coding," DSSD 2010 Decennial Census Memorandum Series #O-A-50, November 14, 2012.

Konicki, Scott (2010a), “2010 Census Coverage Measurement: Block Cluster Sample Extract,” DSSD 2010 Census Coverage Measurement Memorandum Series #2010-C-06R1, October 20, 2010.

Konicki, Scott (2010b), “2010 Census Coverage Measurement: Sample Design – Revised,” DSSD 2010 Census Coverage Measurement Memorandum Series #2010-C-13R1, December 28, 2010.

Konicki, Scott (2011), “2010 Census Coverage Measurement: Sample Design File Documentation – Final Layout,” DSSD 2010 Census Coverage Measurement Memorandum Series #2010-C-26, December 8, 2011.

Lynch, Maureen P. (2009), “Interface Control Document for the Data Exchanges between DSSD CEE and DSPO PBO Version 1.00,” DSSD, last accessed July 25, 2012.

Mazur, Christopher and Ellen Wilson, “Housing Characteristics: 2010,” 2010 Census Briefs, http://www.census.gov/prod/cen2010/briefs/c2010b0-07.pdf, issued October 2011, last accessed June 26, 2012.

Miller, Rupert G. (1974), “The Jackknife—A Review,” *Biometrika* Vol. 61, No.1, pp. 1-15, April, 1974, http://www.jstor.org/stable/2334280, last accessed July 25, 2012.

Mulligan, James M. (2011), “2010 Census Coverage Measurement: E-Sample Identification Results,” DSSD 2010 Census Coverage Measurement Memorandum Series #2010-C-25, August 1, 2011.

Owens, Karen (2008), "Customer Requirements for the 2010 Address Canvassing Address Products Delivery," DSSD 2010 Decennial Census Memorandum Series #D-01R2, November 14, 2008.

Pennington, Robin A., Cynthia A. Rothhaas, and Ryan King (2011), “Clarifications for Analysts using the Decennial Response File, Census Unedited File, and Census Edited File”, DSSD 2010 Decennial Census Memorandum Series #G-34, April 15, 2011.

Probst, Diane (2012), “Interface Control Document (ICD) for DSSD CEE to NPC MaRCS File Transfers to Support the Housing Unit Evaluation Operation,” DSSD 2010 Decennial Census Memorandum Series #H-35R1, January 23, 2012.

Ruhnke, Megan (2002), “The Address Listing Operation and Its Impact on the Master Address File,” Census 2000 Evaluation, U.S. Census Bureau, January 30, 2002.

Ruhnke, Megan (2003), "An Assessment of Addresses on the Master Address File "Missing" in the Census or Geocoded to the Wrong Collection Block," Planning, Research, and Evaluation Division, August 19, 2003.

Smith, Damon, Diane F. Barrett, and Michael Beaghen (2003), "Analysis of Deleted and Added Housing Units in Census 2000 Measured by the Accuracy and Coverage Evaluation," Census 2000 Evaluation O.19, Decennial Statistical Studies Division, October 17, 2003.

Spratt, Shermaine E. and Nancy Johnson (2012), "2010 Census Evaluation of Address Frame Accuracy and Quality Before Followup Clerical Matching Procedures for the Geocoding Error Operation," DSSD 2010 Decennial Census Memorandum Series, #O-A-36, June 28, 2012.

Tomaszewski, Christine, G. (2010), "2010 Census Evaluation Study Plan: Evaluation of Address List Maintenance Using Supplemental Data Sources," DSSD 2010 Decennial Census Memorandum Series #O-A-1, March 22, 2010.

Tomaszewski, Christine, G. (*Forthcoming report 2013*), "Evaluation of Address List Maintenance Using Supplemental Data Sources," 2010 Census Program for Evaluations and Experiments #A-06, *Forthcoming report 2013*.

U.S. Census Bureau (2004), Address List Development in Census 2000, Census 2000 Topic Report No. 8, Census 2000 Testing, Experimentation, and Evaluation Program, March 2004.

U.S. Census Bureau Internal Document (2010), "2010 CUF Overview", PSS-DEV-PPCUF-2010SD.wpd, written by Decennial Systems and Processing Office, April 7, 2010

Viehdorfer, Colt (2011), "The Design of the Coverage Measurement Program for the 2010 Census-Revision #1," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-B-07R1, November, 7, 2011.

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," 2010 Census Program for Evaluations and Experiments, #A-01, February 21, 2012.

Wakim, Anne M. and Diane M. Cronkite (2011), "Specifications for the 2010 Census Coverage Measurement Final Housing Unit Before Followup and After Followup Clerical Matching," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-D4-13, September 27, 2011.

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," 2010 Census Program for Evaluations and Experiments, #A-04, June 13, 2012.

Whitford, David C. (2009). "2010 Census Coverage Measurement: Block Cluster Delineation – Revised," DSSD 2010 Census Coverage Measurement Memorandum Series #C-03R1, February 19, 2009.

Woodward, Jeanne and Bonnie Damon, "Housing Characteristics: 2000," Census 2000 Brief, http://www.census.gov/prod/2001pubs/c2kbr01-13.pdf, issued October 2001, last accessed June 26, 2012.

Zhang, Bei (2010), "Customer Requirements for the Census Final Collection and Tabulation Address Products Deliveries," DSSD 2010 Decennial Census Memorandum Series #D-17R1, June 4, 2010.

Appendix

Table A-1, breaks out the unweighted and weighted add totals by enumeration status across regions. The South contained 5.3 million total add actions -- this was two to three times more add actions than any other region. The South also contained the most correct add actions, 43.7 percent, and added-in-error actions, 7.0 percent. The other three regions contained a more even distribution of all add actions, between 1.4 million and 2.2 million.

Table A-1. 2010 CPEX Address Frame Accuracy and Quality: Status of AC Adds by Region

Status of AC Adds by Region	Unweighted AC Adds	Weighted AC Adds	Weighted Percent (SE[1])
Total AC Adds…….	**40,946**	**10,585,463**	**100.0**
Correct Enumerations……	34,532	8,853,529	83.6 (1.3)
Midwest…….	3,817	1,200,567	11.3 (1.3)
Northeast…….	3,864	1,249,522	11.8 (1.8)
South……	17,122	4,621,334	43.7 (2.4)
West…….	9,729	1,782,106	16.8 (2.6)
Erroneous Enumerations…….	6,414	1,731,934	16.4 (1.3)
Midwest……..	738	342,822	3.2 (1.2)
Northeast……..	915	278,667	2.6 (0.4)
South…….	2,583	739,342	7.0 (0.6)
West……..	2,178	371,103	3.5 (0.4)

[1] Standard Error.
Source: AC Add Analysis File.

Table A-2 explores the distribution of add actions by TEA.

- Mailout/Mailback areas had the highest number of added-in-error actions, at 1.3 million out of 10.6 million add actions. The most records added-in-error were also in Mailout/Mailback areas in 2000.

- Update/Leave areas accounted for only 2.2 million added records and 325 thousand added-in-error actions.

- Update/Enumerate had the fewest add actions and added-in-error units, at 360 thousand total and 44 thousand, respectively.

Table A-2. 2010 CPEX Address Frame Accuracy and Quality: Status of AC Adds by Type of Enumeration Area

Type of Enumeration Area	Unweighted AC Adds	Weighted AC Adds	Weighted Percent (SE[1])
Total AC Adds……	**40,946**	**10,585,463**	**100.0**
Correct Enumerations……	34,532	8,853,529	83.6 (1.3)
Mailout/Mailback[2]…….	24,949	6,578,558	62.1 (2.4)
Update/Leave[3]……...	6,551	1,958,737	18.5 (1.6)
Update/Enumerate[4]……..	3,032	316,235	3.0 (0.7)
Erroneous Enumerations…….	6,414	1,731,934	16.4 (1.3)
Mailout/Mailback……..	4,685	1,362,468	12.9 (1.3)
Update/Leave……..	997	325,050	3.1 (0.3)
Update/Enumerate…….	732	44,416	0.4 (0.1)

[1] Standard Error.
[2] Includes the Mailout/Mailback (TEA=1) and Military (TEA=6).
[3] Includes the Update/Leave (TEA=2) and Urban Update/Leave (TEA=7).
[4] Includes the Remote Update/Enumerate (TEA=3) and Update/Enumerate (TEA=5).
Source: AC Add Analysis File.

In Table A-3 the predominant ACT is assigned to a collection block prior to AC so blocks thought to be empty are not assigned an address characteristic for this evaluation. Blocks with no housing units and non-residential blocks are combined into one category. The remaining categories are: all City-Style, Mixed, (this includes blocks that are at least 70 percent City-Style) and finally blocks that contain zero to less than 70 percent City-Style records. This includes P.O. boxes and location descriptions.

- The majority of added-in-error actions were in all City-Style blocks, at 9.6 percent of all adds.

Table A-3. 2010 CPEX Address Frame Accuracy and Quality: Status of AC Adds by Collection Block Type of Addresses

Status of AC Adds by Collection Block Type of Addresses	Unweighted AC Adds	Weighted AC Adds	Weighted Percent (SE[1])
Total AC Adds……	**40,946**	**10,585,463**	**100.0**
Correct Adds……..……	34,532	8,853,529	83.6 (1.3)
All City-Style…….	16,026	3,775,729	35.7 (2.5)
Mixed and Non-City-Style……..…	10,699	2,611,121	24.7 (1.7)
Non-City-Style……	4,738	1,162,862	11.0 (1.1)
Unknown/Nonresidential……..….	3,069	1,303,817	12.3 (3.0)
Erroneous Adds……..	6,414	1,731,934	16.4 (1.3)
All-City-Style…….	3,450	1,019,939	9.6 (1.3)
Mixed and Non-City-Style……..……	1,827	1,303,817	4.1 (0.4)
Non-City-Style……	793	186,986	1.8 (0.3)
Unknown/Nonresidential…….…….	344	91,830	0.9 (0.3)

[1] Standard Error.
Source: AC Add Analysis File.

Table A-4 breaks down the number of delete and duplicate actions by classification of collection block ACT. Even though a record may only have an official delete action, it could still represent a duplicate of another record. Similarly, a duplicate action generally resulted in a record being deleted from the UCM. Because delete and duplicate categories were not mutually exclusive, all delete/duplicate numbers are presented in aggregate.

- The vast majority of delete/duplicate actions were deletes, at 14.8 million, or 81 percent of all delete/duplicate actions.

- The 3.6 million true duplicates were determined by using the GQV file unit status and AC action variables.

Table A-4. 2010 CPEX Address Frame Accuracy and Quality: Count of AC Deletes/Duplicates

AC Delete/Duplicate Status	*2010 AC*		
	Unweighted Count	Weighted Count	Weighted Percent (SE[1])
Total AC Deletes/Duplicates……	**30,476**	**18,445,131**	**100.0**
Delete Actions……	24,641	14,853,647	81.0 (1.3)
Duplicate Actions……...	5,835	3,591,484	19.5 (1.3)

[1] Standard Error.
Source: AC Delete/Duplicate Analysis File.

Table A-5 shows the distribution of delete actions by TEA.

- The majority of delete actions occurred in Mailout/Mailback areas at 14.8 million or 80.6 percent of all delete actions.
- The majority of deleted-in-error actions also occurred in Mailout/Mailback areas, at 87 percent, or 684k thousand records out of 786k thousand deleted-in-error records.
- There were not enough deleted-in-error cases in in Update/Leave and Update/Enumerate areas to produce weighted counts.

Table A-5. 2010 CPEX Address Frame Accuracy and Quality: Status of AC Deletes/Duplicate by Type of Enumeration Area

Status/Type of Enumeration Area	Unweighted AC Deletes	Weighted AC Deletes	Weighted Percent (SE[1])
Total AC Deletes/Duplicates……...	**30,476**	**18,445,131**	**100.0**
Correctly Deleted…….	29,171	17,658,837	95.7 (0.6)
Mailout/Mailback[2]…….	23,708	14,182,709	76.9 (1.8)
Update/Leave[3]……	4,222	3,094,881	16.8 (1.7)
Update/Enumerate[4]……	1,241	381,247	2.1 (0.6)
Deleted-in-Error…….	1,305	786,294	4.3 (0.6)
Mailout/Mailback[2]…….	1,139	684,833	3.7 (0.6)
Update/Leave[3]……	131	*	*
Update/Enumerate[4]……	35	*	*

Source: AC Delete/Duplicate Analysis File.
[1] Standard Error.
[2] Includes the Mailout/Mailback (TEA=1) and Military (TEA=6).
[3] Includes the Update/Leave (TEA=2) and Urban Update/Leave (TEA=7).
[4] Includes the Remote Update/Enumerate (TEA=3) and Update/Enumerate (TEA=5).
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table A-6 examines the distribution of delete actions by comparing City-Style collection blocks to Non-City-Style blocks.

- Over 80 percent of delete/duplicate actions occurred in City-Style address collection blocks.
- Out of 786 thousand deleted-in-error units, 708 thousand were in City-Style collection blocks.
- Only 129- unweighted deleted-in-error actions were in Non-City-Style blocks, this was not enough cases to produce a reliable weighted estimate.

Table A-6. 2010 CPEX Address Frame Accuracy and Quality: Status of AC Delete/Duplicates by Collection Block Type of Addresses

Collection Block Type of Addresses	Unweighted AC Delete/Duplicates	Weighted AC Delete/Duplicates	Weighted Percent (SE[1])
Total AC Delete/Duplicates……..	**30,476**	**18,445,131**	**100.0**
Correct Deletes……	29,171	17,658,837	95.7 (0.6)
City Style……	23,772	14,234,419	77.2 (1.7)
Non-city-style……..	5,399	3,424,418	18.6 (1.7)
Erroneous Deletes……	1,305	786,294	4.3 (0.6)
City Style……	1,176	708,546	3.8 (0.6)
Non-city-style……..	129	*	*

Source: AC Delete/Duplicate Analysis File.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.
[1] Standard Error.

Table A-7 shows the distribution of CUF Adds by TEA. According to the FHU matching results, the Mailout/Mailback TEA (which includes the Military TEA) had both the highest percentages of correctly added units (64.9 percent) and the highest percentage of units added-in-error (14.2 percent).

Table A-7. 2010 CPEX Address Frame Accuracy and Quality: Status of Census Unedited File Adds by Type of Enumeration Area

Type of Enumeration Area	Unweighted CUF Adds	Weighted CUF Adds	Weighted Percent (SE[1])
Total CUF Adds ……	**4,769**	**3,601,110**	**100.0**
Correct Enumerations (Correctly Added by Census)…….	3,449	2,867,070	79.6 (2.5)
Mailout/Mailback [2]…….	2,421	2,336,069	64.9 (4.2)
Update/Leave[3]…….	673	479,924	13.3 (2.1)
Update/Enumerate[4]…….	355	51,076	1.4 (0.4)
Erroneous Enumerations (Added-in-Error by Census)....	1,320	734,040	20.4 (2.5)
Mailout/Mailback[2]……..	891	510,817	14.2 (2.0)
Update/Leave [3]……	310	195,359	5.4 (0.9)
Update/Enumerate[4]…….	119	*	*

[1] Standard Error.
[2] Includes the Mailout/Mailback (TEA=1) and Military (TEA=6).
[3] Includes the Update/Leave (TEA=2) and Urban Update/Leave (TEA=7).
[4] Includes the Remote Update/Enumerate (TEA=3) and Update/Enumerate (TEA=5).
*Estimates omitted – number of cases is too small.
Source: CUF Adds Analysis Files.

Table A-8 shows the status of the CUF New Adds according to the FHU matching results. As indicated in the table, the census operations correctly added 74.5 percent of the CUF New Adds. The CCM program verified that these units existed on Census Day and were not duplicates of other units. The census operations added-in-error the remaining 25.5 percent of the CUF New Adds. These cases included units that CCM verified to be duplicates, that were not a housing unit on Census Day, or that were geocoded to an incorrect block.

Table A-8. 2010 CPEX Address Frame Accuracy and Quality: Status of Census Unedited File New Adds in Final Housing Unit Results

CUF New Add Status	Unweighted CUF New Adds	Weighted CUF New Adds	Weighted Percent (SE[1])
Total CUF New Adds……	**2,273**	**1,288,754**	**100.0**
Correctly Added……..……..	1,623	960,527	74.5 (2.8)
Added-in-Error……...…	650	328,227	25.5 (2.8)

[1] Standard Error.
Source: CUF Adds Analysis Files.

Table A-9 shows the CCM enumeration status of the CUF New Adds in the FHU matching results.

Table A-9. 2010 CPEX Address Frame Accuracy and Quality: Status of Census Unedited File New Adds in CCM Final Housing Unit Results

CCM Enumeration Status	Unweighted CUF New Adds	Weighted CUF New Adds	Weighted Percent (SE[1])
Total CUF New Adds……	**2,273**	**1,288,754**	**100.0**
Correct Enumerations (Correctly Added by Census)……	1,623	960,527	74.5 (2.8)
Matches/Possible Matches…….	800	429,606	33.3 (3.9)
Census Nonmatches……...	549	372,768	28.9 (5.2)
Geocoding Error within Search Area…….	140	*	*
Unresolved Units……	134	*	*
Erroneous Enumerations (Added-in-Error by Census)……	650	328,227	25.5 (2.8)
Not a Housing Unit……	306	125,548	9.7 (1.3)
Duplicates……...	301	180,464	14.0 (2.1)
Geocoding Error……	43	*	*

[1] Standard Error.
Source: CUF Adds Analysis Files.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table A-10 shows the distribution of CUF Reinstated Adds in the FHU matching results.

Table A-10. 2010 CPEX Address Frame Accuracy and Quality: Status of Census Unedited File Reinstated Adds in CCM Final Housing Unit Results

CUF Reinstated Adds Status	Unweighted CUF Reinstated Adds	Weighted CUF Reinstated Adds	Weighted Percent (SE[1])
Total CUF Reinstated Adds…….	**2,496**	**2,312,356**	**100.0**
Correctly Added……..	1,826	1,906,543	82.5 (3.3)
Added-in-Error…….	670	405,812	17.5 (3.3)

[1] Standard Error.
Source: CUF Adds Analysis Files.

Table A-11 shows the distribution of the CUF Reinstated Adds by the enumeration status in FHU.

Table A-11. 2010 CPEX Address Frame Accuracy and Quality: Status of Census Unedited File Reinstated Adds in CCM Final Housing Unit Results

CCM Enumeration Status	Unweighted CUF Reinstated Adds	Weighted CUF Reinstated Adds	Weighted Percent (SE[1])
Total CUF Reinstated Adds…….	**2,496**	**2,312,356**	**100.0**
Correct Enumerations (Correctly Added by Census)…….	1,826	1,906,543	82.5 (3.3)
Matches……..	730	780,267	33.7 (6.6)
Census Nonmatches……	631	857,700	37.1 (7.4)
Geocoding Error within Search Area….	453	263,900	11.4 (3.0)
Unresolved Units…….	12	*	*
Erroneous Enumerations (Added-in-Error by Census)…….	670	405,812	17.5 (3.3)
Not a Housing Unit…….	117	*	*
Duplicates……	367	247,197	10.7 (2.5)
Geocoding Error…….	186	*	*

[1] Standard Error.
Source: CUF Adds Analysis Files.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table A-12 shows the distribution of CUF Deletes/Duplicates in the FHU matching results by the TEA. According to the FHU, the Mailout/Mailback TEA had the largest percent of CUF Deletes/Duplicates correctly deleted (62.3 percent) and deleted-in-error (20.9 percent).

Table A-12. 2010 CPEX Address Frame Accuracy and Quality: Status of Census Unedited File Deletes/Duplicates in Final Housing Unit Matching by Type of Enumeration Area

Status/Type of Enumeration Area	Unweighted CUF Deletes	Weighted CUF Deletes	Weighted Percent (SE[1])
Total CUF Deletes/Duplicates……	**8,013**	**4,850,528**	**100.0**
Correctly Deleted…….	5,854	3,599,162	74.2 (1.8)
Mailout/Mailback[2]……..	4,290	3,024,227	62.3 (2.4)
Update/Leave[3]……	804	518,681	10.7 (1.3)
Update/Enumerate[4]…….	760	56,254	1.2 (0.2)
Deleted-in-Error……...	2,159	1,251,366	25.8 (1.8)
Mailout/Mailback[2]…….	1,419	1,015,911	20.9 (1.6)
Update/Leave[3]……	254	183,822	3.8 (0.5)
Update/Enumerate[4]…….	486	51,633	1.1 (0.3)

[1] Standard Error.
[2] Includes the Mailout/Mailback (TEA=1) and Military (TEA=6).
[3] Includes the Update/Leave (TEA=2) and Urban Update/Leave (TEA=7).
[4] Includes the Remote Update/Enumerate (TEA=3) and Update/Enumerate (TEA=5).
Source: CUF Delete/Duplicate Analysis File.

Table A-13 displays the geocoding status of records across census regions.

- The South had the most misgeocoded units, at 1.2 million out of all 2 million misgeocoded records. The South also had the most correctly geocoded units at 50 million out of 130 million records. It is important to note that the South had 40 percent of all valid records.

- In contrast, the Northeast had the fewest records overall, at 18.1 percent, but did not have the fewest misgeocoded records.

- The Midwest had the fewest misgeocoded records at 206 thousand.

Table A-13. 2010 CPEX Address Frame Accuracy and Quality: Status of Geocoding by Census Region

Status of Geocoding by Region	Unweighted Records	Weighted Records	Weighted Percent (SE[1])
Total Records……	**478,555**	**132,538,124**	**100.0**
Correctly Geocoded……..	471,836	130,543,009	98.5 (0.1)
Midwest……...	74,899	28,743,504	21.7 (0.7)
Northeast……..	77,658	23,957,852	18.1 (0.6)
South……	185,908	50,448,359	38.1 (1.0)
West…….	133,375	27,394,159	20.7 (0.7)
Erroneously Geocoded……..	6,714	1,992,600	1.5 (0.1)
Midwest……...	454	206,175	0.2 (0.03)
Northeast……..	947	301,871	0.2 (0.03)
South……	3,953	1,191,849	0.9 (0.1)
West…….	1,360	292,706	0.2 (0.04)

[1] Standard Error.
Source: Geocoding Analysis File.

Table A-14 presents the geocoding status of records across TEAs.

Table A-14. 2010 CPEX Address Frame Accuracy and Quality: Status of Geocoding by Type of Enumeration Area

Status/Type of Enumeration Area	Unweighted Records	Weighted Records	Weighted Percent (SE[1])
Total Records……	**478,555**	**132,538,124**	**100.0**
Correctly Geocoded……..	471,836	130,543,009	98.5 (0.1)
Mailout/Mailback[2]……...	428,995	118,382,744	89.3 (0.4)
Update/Leave[3]……..	32,942	10,779,706	8.1 (0.4)
Update/Enumerate[4]……..	9,578	1,274,574	1.0 (0.1)
Unknown TEA……..	321	105,985	0.1 (0.01)
Erroneously Geocoded……	6,714	1,992,600	1.5 (0.1)
Mailout/Mailback[2]……...	5,961	1,752,323	1.3 (0.1)
Update/Leave[3]…….	528	209,750	0.2 (0.03)
Update/Enumerate[4]……..	225	*	*
Unknown TEA……...	0	*	*

[1] Standard Error.
[2] Includes the Mailout/Mailback (TEA=1) and Military (TEA=6).
[3] Includes the Update/Leave (TEA=2) and Urban Update/Leave (TEA=7).
[4] Includes the Remote Update/Enumerate (TEA=3) and Update/Enumerate (TEA=5).
*Estimates omitted – number of cases is too small.
Source: Geocoding Analysis File.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table A-15 presents geocoding error rate broken out by type of structure including multi-unit, single-unit, mobile-home, other, or unknown structure type. It is important to examine the category of multi-unit structures, because if the building is misgeocoded, each unit in the building counted as a separate, additional geocoding error.

- Out of 7.4 million mobile-homes, 139 thousand were misgeocoded, which means the mobile-home geocoding error rate was slightly higher than the national average, at 1.9 percent.

- Out of about 90 million single-units less than one million units were misgeocoded. The single-unit geocoding error rate was only one percent.

Table A-15. 2010 CPEX Address Frame Accuracy and Quality: Status of Geocoding by Type of Structure

Type of Structure	Unweighted Records	Weighted Records	Weighted Percent (SE[1])
Total Records……	**478,555**	**132,538,124**	**100.0**
Correctly Geocoded……	471,836	130,543,009	98.5 (0.1)
Single-Unit…….	252,360	89,546,255	67.6 (0.8)
Multi-Unit……	164,686	28,028,810	21.1 (0.6)
Mobile-Home……..	23,035	7,307,928	5.5 (0.3)
Other……	288	93,858	0.1 (0.01)
Missing/Unknown……..	31,467	5,566,157	4.2 (0.5)
Erroneously Geocoded…….	6,714	1,992,600	1.5 (0.1)
Single-Unit……..	2,511	917,475	0.7 (0.1)
Multi-Unit……	2,473	594,939	0.4 (0.1)
Mobile-Home……..	417	139,929	0.1 (0.02)
Other……	2	*	*
Missing/Unknown……	1,311	339,875	0.3 (0.04)

[1] Standard Error.
Source: Geocoding Analysis File.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table A-16 presents the geocoding status of records across original sources.

Table A-16. 2010 CPEX Address Frame Accuracy and Quality: Status of Geocoding by Original Source

Original Source	Unweighted Records	Weighted Records	Weighted Percent (SE[1])
Total Records……	**478,555**	**132,536,626**	**100.0**
Correctly Geocoded……	471,844	130,544,026	98.5 (0.1)
Pre-2010 Census (Census 2000, Dress Rehearsal, etc.)……	428,997	118,383,050	89.3 (0.4)
DSF……..	135,304	35,454,312	26.8 (0.6)
2010 Address Canvassing……	23,061	5,963,029	3.9 (0.2)
2010 LUCA…….	10,043	2,040,383	1.5 (0.1)
2010 Census Operation(Post AC)…….	3,603	787,567	0.6 (0.1)
Unknown Source…….	415	123,575	0.1 (0.01)
Erroneously Geocoded…….	6,714	1,992,600	1.5 (0.1)
Pre-2010 Census (Census 2000, Dress Rehearsal, etc)……	3,367	1,062,926	0.8 (0.1)
DSF……	2,781	750,773	0.6 (0.1)
2010 Address Canvassing…….	357	102,846	0.1 (0.01)
2010 LUCA……...	93	*	*
2010 Census Operation (Post AC)…….	115	*	*
Unknown Source……...	1	*	*

[1] Standard Error.
Source: Gecocoding Analysis File.
*There was not a sufficient quantity of cases to produce a reliable weighted estimate.

Table A-17 shows the original source for the 2010 Census CUF records flagged as invalid (or deleted). See Table 6 for the original source of the 2010 Census CUF records flagged as valid.

Table A-17. 2010 CPEX Address Frame Accuracy and Quality: Original Source of Census Unedited File Invalid Records

MAF Source by Type of Record	Count	Percent
Total Invalid CUF Records	**4,887,354**	**100.0**
Blank GQHUFLAG[3]	8	<0.1
Housing Unit[1]	**4,887,346**	**100.0**
Unknown MAF Source	6	<0.1
Delivery Sequence File	1,064,644	21.8
Local Update of Census Addresses/LUCA Appeals	645,716	13.2
Operations/MAF Sources before 2010 Census	1,930,370	39.5
INFOCOMM	5,525	0.1
Count Review - HU Misses	8,827	0.2
Address Canvassing	982,846	20.1
Group Quarters Validation	60,672	1.2
Update Leave/UUE/UUL/Remote Alaska/Remote UE/UE	43,901	0.9
New Construction	92,642	1.9
Non-Response Followup/Vacant Delete Check	373	<0.1
Non-ID Operations[2]	51,824	1.1

[1] Includes records with GQHUFLAG equal to blank, 0, 4, and 5. This includes Transient Locations/Units.
[2] Includes Be Counted, Telephone Questionnaire Assistance (TQA) – Fulfillment, TQA – Interview, GQ Enumeration Usual Home Elsewhere (UHE), NRFU Whole Household Usual Residence Elsewhere, and Remote Alaska UHE/Update Enumerate UHE/Remote Update Enumerate UHE.
[3] GQHUFLAG=Group Quarters Housing Unit Flag.
Source: 2010 Census Final Tabulation Master Address File (MAFOP and MAFX) and 2010 Census Unedited File.

Correctly Added Address Canvassing Units

This section describes the CCM match codes that this evaluation defines as correct adds for the 2010 Census AC. In general a unit is a correct add if it represents a valid non-duplicate housing unit on Census Day:

- *Match or M and MS* - Linked CCM and census records that were correctly enumerated and geocoded to the block cluster or the surrounding ring of blocks to the cluster. CCM defined the unit as a valid housing unit on Census Day, this evaluation considers it a correct add.
- *Correct Enumeration or CE* - The census record existed as a housing unit on Census Day and was correctly geocoded in the cluster but the housing unit did not match a unit from CCM's IL.

- *Possible matches or P* - The CCM and/or census record did not contain enough information to assign a match with confidence. The record is a valid housing unit but it may or may not match the added record. This evaluation gives census the benefit of the doubt and considers this a correct add action.
- *Unresolved status or UE, MU, and KE* - The UE match code identifies a census address that does not match a CCM IL housing unit, and there was not enough information on the CCM followup form to confirm the address was a housing unit or to confirm the geocode with certainty. The MU match code identifies a record where the CCM IL and census addresses match, but there was not enough information on the CCM followup form to confirm the address is a housing unit or to confirm the geocode with certainty. The KE match code identifies a census address that does not have enough information to attempt a match- for example, the address was incomplete or missing. To replicate Smith et al. 2003 methodology, this evaluation assumes these records represent valid housing units and are correct adds.
- *Geocoding Error within Search Area or GS* - The census address existed as a housing unit on Census Day and was geocoded to the sample block cluster, but it was actually located in the surrounding blocks to the cluster. CCM coded the unit as correctly enumerated because it was located in the search area. As in Smith et al. 2003 this evaluation considers these to be correct adds because they are geocoding error within the search area.
- *Unconfirmed Duplicate or DF* - The census housing unit was a duplicate identified in AFU that has not been confirmed by the Field Division. The case was identified as a possible duplicate in IHU and was not field confirmed in IHU or FHU. Since it has not been confirmed by field that these records are duplicates this evaluation gives census the benefit of the doubt and assumes these are correct adds.

Address Canvassing Units Added-in-Error

In this section the CCM match codes that indicate a unit added-in-error are defined. If an AC Add had a CCM match code that indicates it was not a valid housing unit on Census Day or that it was actually located outside the ring of blocks that surround the block cluster, then it was added-in-error. The following list contains CCM codes for units added-in-error:

- *Duplicates or DE* - The census address was a duplicate of another census record. The case went to either IHU or FHU for field confirmation. If a unit was field confirmed to be a duplicate of another census record this evaluation considers this an added-in-error action.
- *Not a Housing Unit or EE* - The census address was erroneously listed because the address was not a housing unit in the block cluster or surrounding blocks. The record does not represent a valid housing unit on Census Day, this evaluation considers this a record added-in-error.
- *Matching Addresses are not a Housing Unit or MX and MZ* - The CCM IL and census addresses matched and do not refer to a housing unit on Census Day, but were correctly geocoded in the sample block cluster. For example, the address could have been a merged unit, under construction, future construction, unfit for habitation, demolished or burned down, an empty lot/site in a trailer park, a mobile home that moved, or a non-

existing housing unit, such as a shed, barn, commercial property, or storage unit for non-household goods. Since it is not a housing unit this evaluation considers this to be added-in-error.

- *Geocoding Error or GE and ME* - The GE match code identified a census address that was geocoded to the sample block cluster, but was actually located beyond the surrounding blocks to the sample block cluster. The ME match code identified a case where the CCM IL and census addresses matched, and the unit was located outside the surrounding blocks to the sample block cluster. This evaluation used the same standard as Smith et al. 2003 in determining if a misgeocoded record was added-in-error, if it was beyond the search area is it considered to be added-in-error.

Correctly Deleted Address Canvassing Units

In this section, a correct delete action is defined in terms of its CCM match code. When an AC Delete did not match to any CCM record, it was assumed to be a correct delete. In addition, if this Evaluation matched an AC Delete to a CCM address with a code indicating it was an erroneous enumeration or unresolved, the action was assumed to be a correct delete. The following list describes match codes that were considered to be correct deletes:

- *Non-matching AC Delete/Duplicate*-When an AC Delete had no CCM match code because it was not matched to a CCM record it was assumed to not exist or have insufficient information to be matched. This evaluation gave census the benefit of the doubt and assumed these records were all correct delete actions.
- *Match code of a correct enumeration (CE, M, CI, or P) and linked to a duplicate* – If a deleted record is truly a duplicate of another census record this Evaluation assumes it was correctly deleted. However, in a few instances described here the delete is not a duplicate and it was deleted-in-error. A match code of M identified CCM IL and census addresses that matched. Cases with a P match code were possible matches where the clerical matchers did not have enough information to assign a match with confidence. This evaluation treats P the same as an M. A CE match code identified a census address that existed as a housing unit on Census Day and was correctly geocoded in the sample block cluster. The housing unit was not found in the CCM IL. If a record with any of these match codes was linked to a duplicate (as indicated by the DUPCNT variable), it was assumed the AC delete/duplicate record was the linked duplicate, so the census correctly made the record a delete or duplicate. If the MAFID of the AC Delete and census record were different, the delete was also considered a duplicate. A match code of CI was correctly deleted only if multiple AC Deletes matched to the same CI. This indicated all but one of the AC Deletes was a duplicate.
- *Geocoding Error within Search Area or GS, MS, GI* – A GS match code identified a census address that existed as a housing unit on Census Day and was geocoded to the sample block cluster, but it was actually located in the surrounding blocks to the block cluster. CCM coded the census address as correctly enumerated because it was located in the search area. The MS match code identifies a linked CCM IL and census address, and the unit was located in a surrounding block to the sample block cluster. CCM treated the CCM IL address as geocoding error and the census address as correctly enumerated because the census address was in the search area. However, for this evaluation, units

with these match codes were correctly deleted because CCM found the unit in the surrounding block and census correctly deleted it from the sample block cluster. A GI match code indicated a CCM IL address that existed as a housing unit on Census Day that actually existed outside the sample block cluster.

- *Geocoding Error or GE, and ME* – The GE match code identified a census address that was geocoded to the sample block cluster but was actually located beyond the surrounding blocks to the sample block cluster. The ME match code identified a case where the CCM IL and census addresses matched, and the unit was located outside the surrounding blocks to the sample block cluster. This evaluation considers these to be correct deletes because they are geocoded to the wrong block.
- *Unresolved status or UE, MU, UI, and KE* – The UE match code identified a census address that did not match a CCM IL address, and there was not enough information on the CCM followup form to confirm the address was a housing unit or to confirm the geocode with certainty. The MU match code identified a record where the CCM IL and census addresses matched, and there was not enough information on the CCM followup form to confirm the address was a housing unit or to confirm the geocode with certainty. The KE match code identified an address that did not have enough information to attempt a match – for example, the address was incomplete or missing. The UI match code indicated a CCM IL address where there was not enough information on the CCM followup form to confirm the address was a housing unit or to confirm the geocode with certainty. This evaluation gives census the benefit of the doubt and assumes these were correctly deleted units because they were not valid housing units on Census Day.
- *Duplicates or DE, DF,DA, and DI* – A match code of DE identified a census address that was a possible duplicate of another census address. A DA or DF match code identified a possible census housing unit duplicate that had not been field confirmed. A match code of DI indicated a CCM IL address that was a duplicate of another CCM address. All of these cases are assumed to truly be duplicates and therefore correct deletes.
- *Matching Addresses are not a Housing Unit or MX, MZ* – A match code of MX identified linked CCM IL and census addresses that matched and did not refer to a housing unit on Census Day but were correctly geocoded in the sample block cluster. For example, the address could have been a merged unit, under construction, future construction, unfit for habitation, demolished or burned down, an empty lot/site in a trailer park, a mobile home that moved, or a non-existing housing unit, such as a shed, barn, commercial property or storage unit for non-household goods. The MZ match code was similar to the MX but in the case of the MZ, the IHU matching determined that the linked CCM IL and census addresses had the potential to be valid housing units by the time of Person Interview.
- *Not a Housing Unit or EE, ZI, and XI* – A match code of EE indicated the census address was not a housing unit in the sample block cluster or the surrounding blocks on Census Day. A match code of XI or ZI indicated the CCM IL address was not a valid housing unit on Census Day, the only distinction being the ZI record had the potential to be a valid housing unit by time of the Person Followup. This evaluation considers these to be correct deletes because they were not housing units on Census Day.

Address Canvassing Units Deleted-in-Error

This section defines the CCM match codes that indicate a unit was deleted-in-error. When an AC Delete matched to a valid, non-duplicate, unlinked housing unit, this unit was deleted-in-error. If AC had not deleted the record, it would have remained in the UCM and been matched by CCM.

If a record was deleted by AC, but it was still processed and placed on the UCM, it could have been linked to CCM's IL and resulted in an M match code, or matched a valid unlinked census record, CE. If the MAFID of the deleted or duplicated record was the same MAFID of a record on the UCM, this evaluation considered this a unit deleted-in-error. This evaluation was attempting to measure the error rate of AC Adds and Deletes. If a lister took an erroneous action on a record, it is still an error even if the mistake was corrected for a subset of records during subsequent census processing. The following list describes the CCM match codes, obtained by this evaluations special clerical matching operation, considered to be deleted-in-error:

- *Match or M, P* – A match code of M identified P-sample and census addresses that matched. Cases with a P match code were possible matches where the clerical matchers did not have enough information to assign a match with confidence. These cases were only deleted-in-error if the AC Delete was NOT a duplicate of another census address. In order to be considered deleted-in-error, the census address had to meet two conditions: a) have the same MAFID as the deleted census address and b) have a DUPCNT variable that indicated it was NOT a duplicate.
- *Census Nonmatch or CE* – A CE match code identified a census address that existed as a housing unit on Census Day and was correctly geocoded in the sample block cluster. The housing unit was not found in the CCM IL. As with the aforementioned matches, the delete had to meet two conditions to be considered deleted-in-error: a) have the same MAFID as the CE census address and b) have a DUPCNT variable that indicated it was NOT a duplicate.
- *CCM IL Nonmatch or CI and UI* - A match code of CI identified a CCM IL address that was a valid housing unit missing from the census. If multiple AC Deletes matched to the same CI only one AC Delete was considered deleted-in-error. The UI match code indicated a CCM IL address where there was not enough information on the CCM followup form to confirm the address was a housing unit or to confirm the geocode with certainty. However, these cases were sent to AC Delete Field Followup. If field confirmed they were valid housing units on Census Day, they were treated the same as a match code of CI, and this evaluation classified them as deleted-in-error.

Correctly Added Census Unedited File Units

CUF Adds matching addresses that CCM assigned with either correct enumeration match codes or match codes of unresolved were correctly added units. Correct enumerations were addresses that CCM confirmed as housing units that existed on Census Day. The following describes the codes CCM assigned to these cases (Wakim and Cronkite, 2011):

- *Match or M and MS* – Cases assigned a match code of M were housing units where the P-sample and census addresses match. A match code of MS identified P-sample and census addresses that matched, and the units were located in a surrounding block to the sample

block cluster. For those MS cases, in CCM, the P-sample address was a geocoding error, and the census address was a correct enumeration because it was located in the search area.

- *Correct enumeration or CE* – The E-sample address existed as a housing unit on Census Day and was correctly geocoded in the sample block cluster, but the housing unit did not match a unit in the P sample.
- *Possible matches or P* – A match code of P identified P-sample and census addresses that were possible matches, but there was not enough information to assign a match with confidence.
- *Unresolved status or UE, MU, and KE* – The UE match code identified a census address that did not match a P-sample address, and there was not enough information on the CCM followup form to confirm the address was a housing unit or to confirm the geocode with certainty. The MU match code identified a record where the P-sample and census addresses matched, and there was not enough information on the CCM followup form to confirm the address was a housing unit or to confirm the geocode with certainty. The KE match code identified an address that did not have enough information to even attempt a match – for example, the address was incomplete or missing.
- *Geocoding Error within Search Area or GS* – The E-sample address existed as a housing unit on Census Day and was geocoded to the sample block cluster, but it was actually located in the surrounding blocks to the block cluster. CCM coded the unit as correctly enumerated because it was located in the search area.
- *Unconfirmed Duplicate or DF* – The census housing unit was a duplicate identified in After Followup that had not been field-confirmed. The case was identified as a possible duplicate in FHU. The case did not go to the FHU Field Followup.

Census Unedited File Units Added-in-Error

In this evaluation analysis, units added-in-error were CUF Adds matching units identified in CCM as erroneous enumerations. The erroneous enumerations refer to units added to the census inventory that the CCM FHU operations verified as not existing on April 1, 2010 in the sample block cluster or the surrounding blocks. The CCM assigned the following codes as erroneous enumerations:

- *Duplicates or DE* – The census address was a possible duplicate of another E-sample address.
- *Not a Housing Unit or EE* – The E-sample address was erroneously listed because the address was not a housing unit in the sample block cluster or the surrounding blocks.
- *Matching Addresses are not a Housing Unit or MX* – The P-sample and census addresses matched and did not refer to a housing unit on Census Day but were correctly geocoded in the sample block cluster. For example, the address could be a merged unit, under construction, future construction, unfit for habitation, demolished or burned down, an empty lot/site in a trailer park, a mobile home that moved, or a non-existing housing unit, such as a shed, barn, commercial property, or storage unit for non-household goods.
- *Geocoding Error or GE and ME* – The GE match code identified an E-sample address that was geocoded to the sample block cluster but was actually located beyond the surrounding blocks to the sample block cluster. The ME match code identified a case

where the P-sample and census addresses matched, and the unit was located outside the surrounding blocks to the sample block cluster. As in Smith et al. 2003 this evaluation considers an add to be in error if it is geocoded outside the immediate search area.

Correctly Deleted Census Unedited File Units

The DSSD classified a unit as correctly deleted if CCM coded the unit as an erroneous enumeration, a duplicate, or with an unresolved enumeration status. Below are the CCM match codes counted as correctly deleted:

- *Not a Housing Unit or EE* – The E-sample address was not a housing unit in the sample block cluster or the surrounding blocks on Census Day.
- *Matching Addresses are not a Housing Unit or MX, MZ* – A match code of MX identified linked P-sample and census addresses that matched and did not refer to a housing unit on Census Day, but were correctly geocoded in the sample block cluster. For example, the address could be a merged unit, under construction, future construction, unfit for habitation, demolished or burned down, an empty lot/site in a trailer park, a mobile home that moved, or a non-existing housing unit, such as a shed, barn, commercial property, or storage unit for non-household goods. The MZ match code is similar to the MX, but in the case of the MZ, the IHU matching determined that the linked P-sample and census addresses had the potential to be valid housing units by the time of person followup.
- *Census Duplicates or DE, DF* – A match code of DE identified a census address that was a possible duplicate of another E-sample address. A DF match code identified a possible census housing unit duplicate that had not been field confirmed.
- *Geocoding Error or GE and ME* – The GE match code identified an E-sample address that was geocoded to the sample block cluster but was actually located beyond the surrounding blocks to the sample block cluster. The ME match code identified a case where the P-sample and census addresses match, and the unit was located outside the surrounding blocks to the sample block cluster.
- *Unresolved status or UE, MU, and KE* – The UE match code identified a census address that did not match a P-sample address, and there was not enough information on the CCM followup form to confirm the address was a housing unit or to confirm the geocode with certainty. The MU match code identified a record where the P-sample and census addresses matched, and there was not enough information on the CCM followup form to confirm the address was a housing unit or to confirm the geocode with certainty. The KE match code identified an address that did not have enough information to even attempt a match – for example, the address was incomplete or missing.
- *Geocoding Error within Search Area or GS, MS* – A GS match code identified an E-sample address that existed as a housing unit on Census Day and was geocoded to the sample block cluster, but it was actually located in the surrounding blocks to the block cluster. CCM coded the census address as correctly enumerated because it was located in the search area. The MS match code identified a linked P-sample and census address, and the unit was located in a surrounding block to the sample block cluster. CCM treated the P-sample address as geocoding error and the census address as correctly enumerated because the census address was in the search area. However, for this

evaluation, units with these match codes were correctly deleted because CCM found the unit in the surrounding block and census correctly deleted it from the sample block cluster.

- *Match code of a correct enumeration (CE, M, or P) and linked to a duplicate* - A match code of M identified P-sample and census addresses that matched. Cases with a P match code were possible matches where the clerical matchers did not have enough information to assign a match with confidence. A CE match code identified an E-sample address that existed as a housing unit on Census Day and was correctly geocoded in the sample block cluster. The housing unit was not found in the P sample. If a record with any of these match codes is linked to a duplicate (as indicated by the FHUFINDUPCNT variable), the analysis assumed the CUF delete/duplicate record was the linked duplicate, so the census correctly made the record a delete or duplicate.

Census Unedited File Units Deleted-in-Error

The DSSD classified units as deleted-in-error if CCM operations confirmed the units existed on Census Day (i.e., correct enumerations) and they were not linked to a duplicate. Below are the CCM match codes for units deleted-in-error:

- *Match or M, P* – A match code of M identified P-sample and census addresses that match. Cases with a P match code were possible matches where the clerical matchers did not have enough information to assign a match with confidence.

- *Census Nonmatch or CE* – A CE match code identified an E-sample address that existed as a housing unit on Census Day and was correctly geocoded in the sample block cluster. The housing unit was not found in the P sample.