

Results of the 2010 Census Coverage Measurement Field and Matching Operations

Patricia Sanchez, Anne Wakim, Diane Cronkite
U.S. Census Bureau, Washington, DC 20233¹

Abstract

Census Coverage Measurement for the 2010 Census will provide estimates of net coverage error and components of census coverage for housing units and persons in housing units. The 2010 estimation process uses a dual system methodology in which a sample is independently enumerated and then compared to the census. In support of this process, five field data collection operations and three matching operations were conducted: Independent Listing created a list of housing units in the sample area; Initial Housing Unit Matching of the Independent Listing and the 2010 Census addresses; Initial Housing Unit Followup collected data to resolve issues from the matching (e.g., unresolved unit status); Person Interview created a roster of people in the selected housing units, Person Matching compared the Person Interview roster with the entire 2010 Census roster; Person Followup collected data to resolve issues from the matching (e.g., possible matches); Final Housing Unit Matching updated the earlier results with changes to the Census address list; and Final Housing Unit Followup collected data to resolve issues from the matching. This paper discusses results from each of these operations.

Key Words: 2010 Census, Coverage measurement, data collection, record linkage

1. Census Coverage Measurement Introduction

The purpose of the 2010 Census Coverage Measurement (CCM) program was to evaluate coverage error in the 2010 Census in order to improve future censuses, meaning 2020 and beyond. The CCM was designed to measure the coverage of housing units and persons in housing units in the United States (U.S.) and Puerto Rico. The CCM program excluded remote areas in Alaska. Coverage of group quarters facilities and person residing in those facilities was also out of scope. The CCM provided estimates of the net coverage error and the components of census coverage, including omissions and erroneous enumerations. Since the CCM was an evaluation, its results did not affect the 2010 Census.

The 2010 CCM was a large, complex survey conducted independently of the census. The CCM included five sampling activities, five data collection activities, six matching activities, and separate estimation of the national housing unit coverage and coverage of the U.S. population as of April 1, 2010. This paper will discuss results from the data collection and matching activities conducted in the U.S., excluding Puerto Rico. These activities include:

- CCM Independent Listing
- CCM Initial Housing Unit Matching (Computer and Clerical)

¹ Disclaimer - This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed are those of the author(s) and not necessarily those of the U.S. Census Bureau.

- CCM Initial Housing Unit Followup
- CCM Person Interview
- CCM Person Matching (Computer and Clerical)
- CCM Person Followup
- CCM Final Housing Unit Matching (Computer Processing and Clerical)
- CCM Final Housing Unit Followup

1.1 2010 Census Coverage Measurement Sample Design

The CCM was a multi-phase sample designed to measure net coverage as well as the components of coverage for the housing units and household population in the 2010 Census. The CCM employed dual system estimation methodology and was based on two samples. The Population sample (P sample) provided information about people and housing units missed by the census. The P sample was populated from housing units listed in the CCM and the people rostered in those housing units during the Person Interview (PI). The Enumeration sample (E sample) provided information about erroneous census inclusions. The E sample was populated from housing units included in the census and from census enumerations in those housing units.

The primary CCM sampling unit was a block cluster, which consisted of one or more geographically contiguous census blocks. A stratified sample of block clusters was selected for each state, the District of Columbia, and Puerto Rico. An independent CCM address list was created for each CCM sample block cluster during Independent Listing (IL). Based on the IL, the small block clusters were identified and then subsampled for inclusion in the remaining CCM operations. Census and CCM addresses in 6,148 block clusters were included in Initial Housing Unit (IHU) Matching and Followup operations. Output from those operations provided a list of CCM addresses eligible for the P sample. The P sample consisted of 171,217 housing units and potential housing units² that were included in the PI. For block clusters with fewer than 80 CCM housing units, all units (the complete block cluster) were selected with certainty. For block clusters with 80 housing units or more, a subsample of units was selected. The E sample consisted of 180,528 census housing units selected from the census files in the same block clusters as the P sample.

2. Results

2.1 2010 Census Coverage Measurement Independent Listing

The IL operation built the housing unit address list for further CCM operations and provided data for housing unit estimation. This operation was conducted independently of census address canvassing. CCM addresses from IL were compared to the census addresses during the IHU Matching operations.

During IL, scheduled from August 28, 2009 to December 12, 2009, listers canvassed each block cluster assigned, and listed in paper Independent Listing Books (ILBs) all housing units and potential housing units². Listers marked the location of each unit they listed on paper CCM maps provided to them. Listers also updated the CCM maps by adding and deleting streets based on their observations.

² Potential housing unit included units under construction, empty mobile home sites, etc. that may have become valid housing units at the time of the PI - August 14, 2010 to October 16, 2010.

For reasons of weather-related issues and management of field work, the operation was conducted as planned in three overlapping 6-week periods from August 28, 2009 through December 5, 2009, across the U.S.

Table 1: 2010 CCM Independent Listing Workload

Initial Workload	Block Clusters ¹	Collection Blocks ²	Total Housing Units
U.S. Total	11,835	17,433	834,223
¹ Block Cluster – A small geographic area consisting of a single census block or a group of census blocks. It is the basic unit for data collection by a single CCM lister or other field staff. ² Collection Block – A physical block enumerated as a single geographic area, regardless of any legal or statistical boundaries passing through it. (Note: State, county, American Indian area, and military base boundaries, as recorded in the TIGER® database at the time of assigning numbers to collection blocks, are always block boundaries.)			

2.2 2010 Census Coverage Measurement Initial Housing Unit Matching

During IHU Computer Matching, we compared the housing units listed by CCM with the housing units and group quarters listed by census. After standardizing the CCM and census addresses, the CCM addresses were computer matched in each block cluster with the census addresses in the block cluster and one ring of surrounding blocks.³ The Computer Matching began on January 26, 2010 and ended on February 23, 2010. As a result of computer matching, CCM and census addresses were assigned one of four computer match codes: match, possible match, nonmatch, or duplicate.

During IHU Before Followup (BFU) Clerical Matching, staff reviewed the output from computer matching. The IHU BFU Clerical Matching operation began on 2/16/2010 and ended on 3/26/2010. Staff used computer-assisted clerical matching techniques, along with CCM and census maps, to verify the match codes from computer matching and to search for additional matches as well as additional duplicate records in the CCM and census files. Cases that remained unresolved following this operation were eligible for Initial Housing Unit Followup (IHUFU), a field operation. In IHU After Followup (AFU) Clerical Matching, staff reviewed the completed paper questionnaires from IHUFU to resolve the cases that were sent to followup. The IHU AFU Clerical Matching operation began on 3/29/2010 and ended on 5/19/2010. The result of this operation was a set of files containing match codes for CCM and census addresses in the search areas, where a search area is defined as a CCM sample block cluster and one ring of blocks surrounding the block cluster. Although the IHU matching operations included all census addresses in the CCM search areas, the results provided here are limited to census housing units located within the CCM sample block clusters i.e., the results exclude census housing units in the surrounding blocks and all census group quarters.

Table 2 summarizes the matching results through each stage of matching and clerical review. The BFU clerical review of the computer matching results substantially increased the percentage of matched of CCM and census addresses. On the other hand, the review of the followup data in AFU matching yielded a much smaller increase in the percent of

³ We included addresses in surrounding blocks to allow for possible discrepancies in the assignment of block codes.

matched units. In addition to finding matches, a major goal of CCM was to detect duplication in the census. During each of the clerical review stages, CCM did find additional census duplicates. The percentage of census housing units identified as duplicates increased from 0.15 percent in computer matching to 0.34 percent in BFU Clerical Matching and up to 1.03 percent in AFU Clerical Matching.⁴

Table 2: The 2010 CCM Initial Housing Unit Matching Operation Matching Results Following Each Stage (Percent of Total)						
	CCM Housing Units 489,592 Housing Units			Census Housing Units in the CCM Sample Areas 484,150 Housing Units		
	Computer Matching	Clerical Matching		Computer Matching	Clerical Matching	
		Before Followup	After Followup		Before Followup	After Followup
Matches	66.85	89.87	92.91	66.29	89.10	92.03
Possible Matches	14.07	3.62	n/a	13.37	3.51	n/a
Nonmatches	18.83	6.42	3.92	20.19	7.04	3.82
Duplicates	0.26	0.09	0.09	0.15	0.34	1.03
Not a Housing Unit ⁵	n/a	<0.01	3.09	n/a	n/a	3.13

In addition to the tabulations by match code, results are tabulated by the housing unit status of the CCM addresses and the enumeration status of the census addresses.

Each CCM unit from IL was classified as either a housing unit, potential housing unit, not a housing unit, duplicate, geocoding error, or unresolved, based on the match code assigned to the unit at the end of AFU Clerical Matching. A CCM unit was a housing unit unless classified as one of the following based on the clerical review of the IHUFU questionnaires.

- Potential housing unit: The CCM unit did not exist as a housing unit at the time of IHUFU, but it had the potential to become a housing unit by the time of the PI. The unit could be under construction, future construction, unfit for habitation, demolished or burned down, or an empty mobile home site. Potential housing units were given a chance to be selected for the PI.
- Not a housing unit: The CCM unit did not exist as a housing unit at the time of IHUFU and did not have the potential to become a housing unit by the time of the PI. It could have been a business, a barn, or a unit that was merged with another housing unit. A CCM unit that was determined to be a group quarters at the time of IHUFU is also classified as not a housing unit. Units classified as “not a housing unit” were not included in the PI.

⁴ It is possible for a census unit, listed within the CCM sample block cluster and coded as a match, possible match or nonmatch, to have a duplicate, but the duplicate record was geocoded outside of the block cluster. Those duplicate records are not included in these results.

⁵ In Table 2 only, this category also includes those CCM units with the potential to be a housing unit. This category applies mainly to AFU results. During BFU, however, errors were detected in a few CCM units that precluded their use in future operations. These units were also classified as “not a housing unit.”

- Geocoding error: The CCM unit was actually located in a block outside of the sample block cluster in which it was listed.
- Duplicate: The CCM unit was the same as another CCM unit listed in the sample block cluster.
- Unresolved: There was not enough information from IHUFU to confirm that the unit was a housing unit or potential housing unit or to confirm that it was located in the sample block cluster.

Each census unit listed by census as a housing unit within the CCM sample block clusters was given an enumeration status of correct enumeration, erroneous enumeration, duplicate, geocoding error, or unresolved, based on the match code assigned to the unit at the end of AFU Clerical Matching. A census unit was classified as a correct enumeration unless classified as one the following based on the results of IHUFU.

- Geocoding Error: Census listed the unit in the sample block cluster, but it was actually located in a block beyond the sample block cluster and its surrounding blocks. (Note: A census unit located in a surrounding block but geocoded in the sample block cluster is not considered a geocoding error.)
- Erroneous Enumeration: The census unit did not exist as a housing unit at the time of IHUFU. It could have been under construction, future construction, unfit for habitation, demolished or burned down, an empty mobile home site, a business, a structure used for the storage of non-household goods, or a unit that was merged with another housing unit. This also includes census units that were determined to be a group quarters at the time of IHUFU.
- Duplicate: The census unit was the same as another census unit listed in the sample block cluster or its surrounding blocks.
- Unresolved: There was not enough information from IHUFU to confirm that the census unit was a housing unit or to confirm that it was located in the sample block cluster or its surrounding blocks.

A summary of the housing unit status for CCM and enumeration status for census addresses at the end of IHU operations is given below. The percent of census units classified as correct enumerations is very high, as was the percent of CCM units classified as housing units. Both are over 95 percent.

Table 3: The 2010 CCM Initial Housing Unit Operations Housing Unit/Enumeration Status Final Results (Percent of Total)			
CCM Housing Units 489,592 Housing Units		Census Housing Units in the CCM Sample Areas 484,150 Housing Units	
Housing Unit	95.49	Correct Enumeration	95.61
Potential Housing Unit	2.07	Erroneous Enumeration	3.07
Not a Housing Unit	1.02		
Unresolved	0.16	Unresolved Enumeration	0.24
Duplicate	0.09	Duplicate	1.03
Geocoding Error	1.18	Geocoding Error	0.06

2.3 2010 Census Coverage Measurement Initial Housing Unit Followup

The IHUFU was conducted from February 23, 2010 to May 10, 2011. Interviewers went to specific households to gather information needed to resolve discrepancies or uncertainties remaining after BFU Clerical Matching. The data from IHUFU were used to complete the coding of CCM and census addresses in AFU Clerical Matching. Information was collected to 1) determine if there was a housing unit at the address, 2) determine the correct block where the address was located, 3) determine whether or not units identified as possible matches really were the same, and 4) determine whether or not units identified as possible duplicates really were the same. Interviewers were also instructed to refer to a reference list of addresses, unique for each block cluster, to determine if the address being followed up could be the same as any CCM or census addresses on the list. The IHUFU data collection forms were created via Docuprint technology. The questions included for each followup case varied depending upon the reason the case was being sent to followup.

As an example, an unlinked address listed by census in one of the CCM sample block clusters would have been sent to followup to determine why it was unlinked. Was it listed in wrong block? Was it a really a housing unit? Was it just missed by the IL? Each IHUFU form was tailored to a specific type of case needing followup. For example, there was a special form for a possible CCM/Census match with two census addresses that may be duplicates. That one form was used to collect data for four specific addresses. Twenty-nine different form types were used to collect data.

Table 4: Number of Block Clusters in Initial Housing Unit Followup					
	Total	Block Clusters			
		Requiring Followup		Not Requiring Followup	
		Count	Percent	Count	Percent
U.S. Total	6,148	4,682	76.15	1,466	23.85

Of the 6,148⁶ sample block clusters within the U.S., 4,682 contained cases that required followup. Each case corresponded to one form to be completed in the field, but may have involved multiple addresses. There were a total of 101,210 cases that went to followup involving 141,830 addresses. The majority of those cases involved only one address: 32.68 percent involved a nonmatched census addresses and 31.03 percent involved a

⁶ The number of sample block clusters was reduced from 12,364 in IL to 6,416 in subsequent operations, due to planned subsampling and as part of an initiative to reduce the number of sample block clusters and focus resources on the reduction of non-sampling error. Of the 6,416 block clusters remaining in sample 6,148 were located within the U.S. (excluding Puerto Rico).

nonmatched CCM addresses. A large portion of the cases (15.16 percent) were for possible matches involving two addresses. Other types of cases sent to followup included duplicate housing units, possible group quarters, possible matches in the surrounding blocks, and addresses that needed status updates (housing unit, group quarters, not a housing unit, etc.).

Table 5: Number of Addresses in Initial Housing Unit Followup

	Count	Percent of Total
Total U.S. Addresses	141,830	79.85
CCM Addresses	67,512	47.60
Census Address in Cluster	62,094	43.78
Census Group Quarters or Other Living Quarters in Cluster	243	0.17
Census Housing Unit in Surrounding Block	11,981	8.45

2.4 2010 Census Coverage Measurement Person Interview

The IHU operations resulted in the list of valid, unique CCM housing units located within the CCM sample block clusters. From this address list a set of eligible housing units was determined for the P sample selection to be included in the subsequent CCM person and Final Housing Unit (FHU) operations.

The PI was conducted from August 14, 2010 through October 4, 2010. The goal of the PI was to obtain a roster and information to determine where each resident of the sample housing unit should have been counted on Census Day (April 1, 2010). This included nonmovers and people who may have moved into the selected housing unit since Census Day, known as in-movers. In addition, the PI collected information about persons who moved out of the sample housing unit between Census Day and the time of the PI interview (out-movers). The demographic information collected for each person included name, sex, age, date of birth, race, relationship, and Hispanic origin. It also collected information to determine where each sample unit current resident was living on Census Day and the new address for each out-mover. The PI also collected information about any other alternate addresses where any of the people listed may have been counted in the census. For each alternate address, PI collected other information that could be used to determine the location (ideally a census collection block) of the alternate address, including nearby landmarks, cross streets, and neighbors. The PI was conducted by personal visit interviewers using a computer-assisted data collection instrument on a laptop computer. A telephone wording path was also provided to support instances when respondents requested an interview via telephone. In addition, there was a Spanish translation available. The final data from the PI fed into the matching operation of the census persons enumerated in the CCM sample areas and then into the final estimation of person coverage for the 2010 Census.

The PI workload consisted of 178,765 sampled housing units. Overall, the final response rate for production was 98.38 percent.

The PI was designed to list all people who lived or stayed at the sample address, including people tenuously attached to the address. To encourage the respondent to list

anyone that stayed at the address, in addition to the original roster, there were also four probes that asked about populations that the respondent may not initially consider as part of their household. Once the roster was collected, the PI instrument also collected any individual outmovers who have left the sample address since around Census Day. If the entire household had left (e.g., the house is currently vacant, not a housing unit, or has a whole household that was not staying at the sample address on Census Day), then a separate roster was collected of those people who have moved out. This allowed us to match to any possible person who could have been listed in the Census at this address.

As expected, most people were added in the main roster, but additional people were added through four probes and two outmover rosters for a total of 405,131 people. The average household size for complete interviews was 2.69 people.

In order to determine if a person is a census erroneous enumeration or a duplicate, PI needed to collect alternate addresses where people could have been counted on Census Day other than the sample address. Similar to collecting a roster, the interview asked about different types of alternate addresses to make sure the respondent had provided all the possible addresses where each person could have been counted, such as a college address, a relative's address, as in shared custody situations, or second homes. Alternate addresses also included the Census Day addresses for in-movers.

Overall, 58,362 alternate addresses were collected in the PI. Among all PI cases, 23.15 percent had at least one alternate address and 18.80 percent of all people collected reported an alternate address. For cases that had an alternate address reported, the average number of alternate addresses within a housing unit was 1.40 addresses with the maximum number of unique addresses collected in a single case being 18. For persons with an alternate address, the average was 1.10 addresses per person with the maximum of five unique alternate addresses per person. The average number of people connected to an address is 1.45 with one address reported actually being connected to 30 people.

2.5 2010 Census Coverage Measurement Person Matching

The results of Person Matching were used to develop the CCM estimates of person coverage in the census. Person computer matching compared people rostered in the PI and E-sample people to census enumerations throughout the country⁷ to identify linked PI and census pairs and to identify E-sample people with duplicate census records. Computer matching linked PI people to other PI records within the same sample block cluster to identify PI duplicates. Computer matching used targeted search areas to identify census enumerations that matched or possibly matched to PI people. Targeted search areas were developed based on the geographic areas around the PI person's sample address and alternate address(es). The geographic search area was defined as the block cluster containing the block where the address was located plus one ring of blocks surrounding the block cluster (i.e., surrounding blocks). Using various sources of geographic and address data available at the Census Bureau, an automated "geocoding" procedure attempted to determine the block where the alternate address was located or perhaps even identify a specific address on the Census Bureau's Master Address File that corresponded to the alternate address. In addition, automated residence status coding processed the PI data to determine if the rostered person was a nonmover, inmover, outmover, or out-of-scope of the CCM. People were then classified as P-sample or non P-sample based on their residence status codes. For example, a person who lived in a group quarters on Census Day was out-of-scope and therefore given a non P-sample status.

Clerical staff reviewed the outputs from computer matching and the automated procedures. In addition to verifying the linked records, they also searched for new links between PI and census persons, and searched for any additional PI and census duplicates. Each PI person and each census person in a CCM sample block cluster or surrounding block was coded as a match, possible match, duplicate, possible duplicate or nonmatch. Staff also performed clerical geocoding of alternate addresses and clerical residence status coding of PI people, assigning codes that could not be assigned by the computer and correcting any errors. The results of clerical geocoding, together with the automated geocoding results, were used to identify the search areas needed for clerical matching. Each E-sample person was classified as either as a correct or erroneous enumeration in the census, or was left unresolved due to lack of information. If the PI person or E-sample person needed to go PFU to resolve any discrepancies, clerical staff reviewed the collected PFU data to attempt to resolve the coding of the person during AFU Clerical Matching. The additional alternate addresses collected in PFU were also clerically geocoded. As a result of the automated and clerical geocoding procedures, search areas were identified for 79.19 percent of the 51,474 alternate addresses collected in PI and 75.44 percent of the 27,059 alternate addresses collected in PFU.

Based on the results from automated and clerical residence status coding, of the 392,711 PI people rostered in CCM IL housing units, 92.51 percent were in the P sample, including nonmovers (82.89 percent), inmovers from other inscope addresses (6.85 percent), certain outmovers (0.36 percent), and people who could not be identified as either nonmovers, inmovers or outmovers (2.42 percent). Outmovers were classified as P-sample or non P-sample dependent on whether or not they had a statistical chance to be rostered at their new address as an inmover. If the new address was out-of-scope of the CCM (e.g., a group quarters or unit located outside of the country), the person was

⁷ Person matching was done separately within the U.S. and within Puerto Rico. PI people rostered in the U.S. and E-sample people could only be matched to census enumerations in the U.S. The results given here are for the U.S. only.

counted as a P-sample outmover, otherwise the person was counted as a non P-sample outmover.

As a result of computer and clerical matching, each PI person and census person in a CCM sample block cluster or surrounding block was coded as a match, possible match, duplicate, possible duplicate or nonmatch. In Table 6, we focus on the matching results for P-sample people. Note that the number of P-sample people differs by stage of operation. During the clerical review of the PI and PFU data, PI people may be reclassified from P-sample to non P-sample and vice versa. The percent of P-sample people who matched increased from 84.55 percent after computer matching to 90.50 percent after BFU clerical matching. After the review of PFU data, 91.49 percent of the P-sample people were matches. The percent of duplicates increased after each clerical operation, whereas the percent of possible duplicates and possible matches decreased.

Table 6: The 2010 CCM Person Matching Operations P-sample Matching Results Following Each Stage			
	P-Sample People		
	Computer Matching	Before Followup	After Followup
Total Count for the U.S.	370,853	370,389	363,290
Percent of Total	100.00	100.00	100.00
Matches	84.55	90.50	91.49
Possible Matches	1.42	0.32	0.09
Nonmatches	13.57	8.66	7.89
Duplicates	0.43	0.51	0.53
Possible Duplicates	0.03	0.01	0.00

As mentioned earlier, the E sample was designed to provide estimates of correct and erroneous enumerations in the census. After all matching and clerical review of PI and PFU data, each E-sample person was classified as a correct enumeration, erroneous enumeration, or was left unresolved due to the lack of data. For matched E-sample people, enumeration status was based on the residency status of the linked PI person. Out of 383,537 E-sample people, 89.52 percent were correct enumerations, 4.48 percent were erroneous enumerations, and 6.00 percent were unresolved. Erroneous E-sample enumerations include people that should not have been enumerated anywhere in the census (e.g. fictitious people and people born after Census Day), duplicate enumerations, and people erroneously enumerated within the sample search area who should have been counted elsewhere. Most of the erroneous E-sample enumerations (66.06 percent) were due to duplication; another 29.80 percent should have been counted at another residence located elsewhere in the country.

2.6 2010 Census Coverage Measurement Person Followup

During PFU, interviewers contacted cases identified in the Person BFU Clerical Matching operation as requiring additional information in order to resolve Census Day residence status, enumeration status, match status, or person duplication. PFU ran from January 28, 2011 through March 26, 2011. The PFU workload was identified from the Person BFU Matching activities and there were 56,706 cases which required an interview about one or more people living at a PI address and/or a census address. The PFU data collection was done using paper forms created via Docuprint technology. The questions

included for each followup case and person varied depending upon the reason the person was being sent to followup.

Of the 6,148 U.S. block clusters included in the PI workload, 5,417 (88.1 percent) block clusters had housing units that were included in PFU. There were 108,440 total people selected for followup, 33,639 CCM people and 74,801 census people. There was an average of 1.9 followup people per case.

Table 7: Person Followup Workload by Block Cluster, Cases, and Persons				
Workload	Block Clusters	Cases	CCM Persons	Census Persons
	Count	Count	Count	Count
U.S. Total	5,417	56,706	33,639	74,801

In order to be selected for PFU, a person must have been considered as having sufficient information for followup, which was defined as two characters in the first name and middle initial combined, and two characters in the last name. Insufficient people were excluded because it is difficult for field personnel to ask questions without sufficient name information.

The largest category of P-sample people selected for followup had unresolved residence status (49.81 percent) while the majority of E-sample people selected for followup were nonmatched people (85.36 percent). Additional reasons a person could be sent for followup include things like possible matches and possible duplicates.

2.7 2010 Census Coverage Measurement Final Housing Unit Matching

The FHU operations provided updated files of P-sample and E-sample housing units for development of the CCM estimates of housing unit coverage in the census. Census units may have been added to or deleted from the CCM sample block clusters and surrounding blocks, when comparing the final census file to that used during IHU matching. Census records that were added became available for matching in FHU. Census records that were deleted may have caused breakage in the links made during IHU. The newly unlinked records then became available for matching. Also, updates were made to reflect housing unit status on Census Day. Was the unit a housing unit on Census Day? Recall that the IHU operations were concerned with capturing addresses that could have been housing units at the time of the PI interview. During the PI interview, a question was asked to determine if the address existed as a housing unit on Census Day, and if it was vacant or occupied.

A computer processing operation was performed to prepare the data for FHU clerical matching. Information from the previous IHU and person data collection and matching activities was used to process the P-sample and final census data. As a result, match codes were updated and records were flagged if a clerical review was needed. During the BFU clerical review, staff searched for new matches and new duplicates to the P-sample and E-sample records, in addition to the matches and duplicates found during IHU. Staff also attempted to determine Census Day housing unit status, if the computer was unable to do so. If needed, units were sent to followup to resolve discrepancies or uncertainties, and to confirm possible matches and duplicates, as described in the next section. Data obtained from followup were reviewed during AFU clerical matching and units were recoded as needed. Some units remained unresolved, if CCM was unable to collect the

necessary data to determine the units' match status, Census Day housing unit status, or where the units should have been counted.

The FHU BFU Clerical Matching operation began on 4/27/2011 and ended on 5/24/2011. The FHU AFU Clerical Matching operation began on 5/27/2011 and ended on 6/22/2011. A summary of the matching results upon completion of each of these operations is provided in Table 8. Note that the "Not a Housing Unit" category includes matched and nonmatched units that were not housing units on Census Day or were geocoding errors.

Table 8: The 2010 CCM Final Housing Unit Operation Matching Results Following Each Stage (Percent of Total)				
	P Sample 171,217 Housing Units		E Sample 180,528 Housing Units	
	Before Followup	After Followup	Before Followup	After Followup
Matches	94.78	95.04	88.17	88.40
Possible Matches	0.07	0.01	0.05	0.00
Nonmatches	2.59	2.47	6.32	5.38
Duplicates	0.01	0.01	1.33	1.59
Not a Housing Unit	2.54	2.47	4.13	4.62

Each E-sample unit was given an enumeration status of correct enumeration, erroneous enumeration, duplicate, geocoding error, or unresolved based on the match code assigned to the unit at the end of AFU Clerical Matching. Similarly, each P-sample unit was classified as a housing unit, not a housing unit, duplicate, geocoding error or unresolved. These are the same categories included in the IHU operation, with the exception of possible housing units. In IHU, housing units listed in IL could be classified as possible housing units, i.e., they had the potential to become housing units on Census Day. At the time of FHU, CCM should have had the information needed to determine the unit's actual status on Census Day and recode the P-sample unit as a housing unit or not a housing unit. If that was not the case, the P-sample unit was coded as unresolved. The vast majority of P-sample units (97.43 percent) and E-sample units (93.62 percent) were classified as housing units and correct enumerations, respectively, as shown in Table 9.

Table 9: The 2010 Census Coverage Measurement Final Housing Unit Operation Housing Unit/Enumeration Status upon Completion of All Clerical Matching			
P-Sample		E-sample	
Total U.S. count	171,217	Total U.S. Count	180,528
Percent of Total	100.00	Percent of Total	100.00
Housing Unit	97.43	Correct Enumeration	93.62
Not a Housing Unit	2.47	Erroneous Enumeration	4.41
Duplicate	0.01	Duplicate	1.59
Geocoding Error	0.05	Geocoding Error	0.21
Unresolved	0.04	Unresolved	0.17

2.8 2010 Census Coverage Measurement Final Housing Unit Followup

The FHUFU was conducted from May 9, 2011 to June 13, 2011. Interviewers gathered information needed to resolve discrepancies or uncertainties remaining after BFU clerical matching, for specific addresses. The data from FHUFU were used to complete the coding of P-sample and census addresses in AFU clerical matching. Information was collected to 1) determine if there was a housing unit at the address on Census Day, 2)

determine the correct block where the address is located, 3) determine whether or not units identified as possible matches really were the same, and 4) determine whether or not units identified as possible duplicates really were the same. Interviewers were also instructed to refer to a reference list of addresses, unique for each block cluster, to determine if the address being followed up could be the same as any CCM or census addresses on the list. Twenty-three different form types were used during FHUFU, each with questions tailored for a specific type of followup case. For example, there was a unique form type if the case involved a possible CCM/Census match with two census addresses that may be duplicates. That particular type of case would have involved four different addresses on the same form.

Care was taken to reduce the number of cases that went to both IHUFU and FHUFU. For example, nonmatched P-sample or E-sample cases that had already gone to followup during IHU operations were not selected for followup in FHU operations.

Of the 6,148 block clusters in the U.S., 1,441 contained cases that required person followup. Each case corresponded to one form to be completed in the field, but may have involved multiple addresses. There were a total of 5,153 cases that went to followup consisting of 8,929 addresses. Almost half (49.50 percent) of the cases were nonmatched census address that had been added to the CCM sample block clusters after IHU matching. Many other cases (20.73 percent) involved a census duplicate to a matched census record.

3. Conclusions

The 2010 CCM is a large, complex survey conducted independently of the census. The CCM includes sampling activities, data collection activities, matching activities, and separate estimation of the national housing unit coverage and coverage of the U.S. population as of April 1, 2010. The focus of this paper has been on the data collection and matching activities which included the independent listing of housing units and the rostering of people in those housing units, as well as the matching of housing units and people to the census records. Successful operations were crucial to ensure accurate data for estimation.

The CCM provided data used to evaluate the 2010 Census and produce estimates of census coverage for the populations of housing units and people living in those housing units. Estimates of net coverage as well as the components of coverage were produced using the CCM results for P-sample and E-sample housing units and people. The estimated components of coverage include the number of correct and erroneous enumerations in the census and the number of omissions, for both people and housing units. Estimates of net census coverage and components of coverage were produced at the national level and at various demographic and geographic levels.⁸

For the first time in the Census Bureau history, and in the statistical community, the CCM produced estimates of the components of census coverage. This provided a wealth of data to study where specific errors occurred in the census to be able to research alternative ways of correcting these in future censuses. In addition, the assessments which were summarized in the paper also provide indications and lessons learned about the 2010 CCM operations. These should also allow the Census Bureau to further enhance the CCM operations if similar methods are used for the next census. Major recommendations

⁸ Separate estimates were produced for Puerto Rico.

include simplification of the coding system used for clerical matching operations to rely more on the computer logic to assign the match and duplicate codes based on simple questions presented to the matchers. [More information on all of the operations as well as recommendations for the measuring the coverage of the 2020 Census can be found in the operational assessment documents listed in the References.]

Acknowledgements

The authors would like to thank the following people for their contributions to this paper: Aaron Cantu, Graciela Contreras, Kyra Linse, John Jones, Susanne Johnson, Lora Rosenberger, Gia Donnalley, and Magdalena Ramos.

References

DSSD 2010 CENSUS COVERAGE MEASUREMENT MEMORANDUM SERIES
#2010-I-16 - Assessment for the 2010 Census Coverage Measurement Person Interview Operation

DSSD 2010 CENSUS COVERAGE MEASUREMENT MEMORANDUM SERIES
#2010-I-14R1 - Assessment for the 2010 Census Coverage Measurement Initial Housing Unit Independent Listing, Matching, and Followup Operations – Reissue

DSSD 2010 CENSUS COVERAGE MEASUREMENT MEMORANDUM SERIES
#2010-I-15 - Assessment for the 2010 Census Coverage Measurement Final Housing Unit Matching and Followup Operations

DSSD 2010 CENSUS COVERAGE MEASUREMENT MEMORANDUM SERIES
#2010-I-24 - Assessment for the 2010 Census Coverage Measurement Person Matching and Followup Operations

DSSD 2010 CENSUS COVERAGE MEASUREMENT MEMORANDUM SERIES
#2010-C-27 - 2010 Census Coverage Measurement: Operational Summary of the Sample Design

DSSD 2010 CENSUS COVERAGE MEASUREMENT MEMORANDUM SERIES
#2010-B-07-R1 - The Design of the Coverage Measurement Program for the 2010 Census - REVISION #1